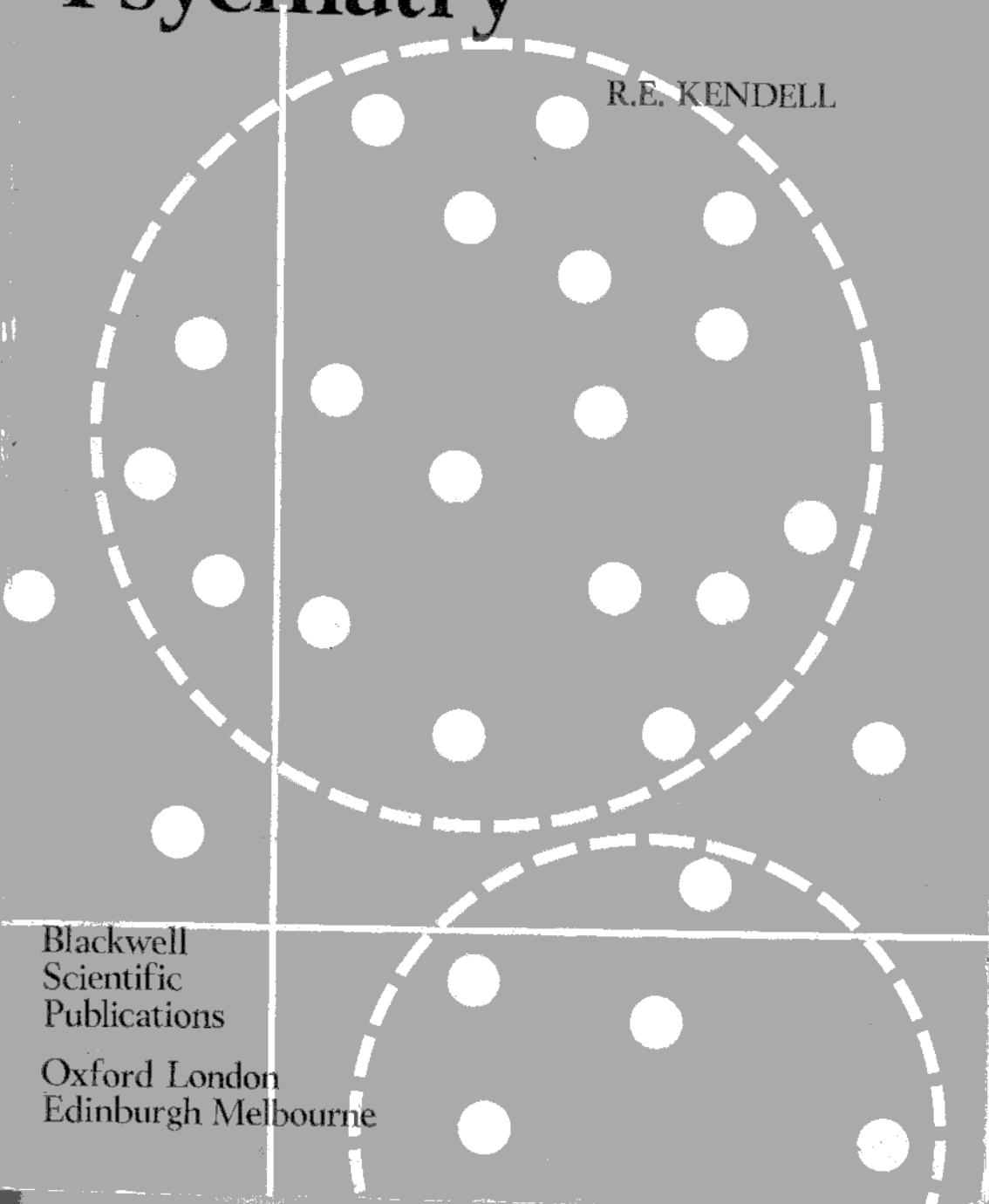


The Role of Diagnosis in Psychiatry

R.E. KENDELL



Blackwell
Scientific
Publications

Oxford London
Edinburgh Melbourne

THE ROLE OF
DIAGNOSIS
IN PSYCHIATRY

THE ROLE OF DIAGNOSIS IN PSYCHIATRY

R. E. KENDELL

M.D., F.R.C.P., M.R.C. Psych.
Professor of Psychiatry
University of Edinburgh

BLACKWELL SCIENTIFIC PUBLICATIONS
OXFORD LONDON EDINBURGH MELBOURNE

© 1975 Blackwell Scientific Publications
Osney Mead, Oxford,
85 Marylebone High Street, London W1,
9 Forrest Road, Edinburgh,
P.O. Box 9, North Balwyn, Victoria, Australia.

All rights reserved. No part of this publication may be reproduced, stored in retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the copyright owner.

ISBN 0 632 00701 X

First published 1975

Distributed in the United States of America by
J. B. Lippincott Company, Philadelphia and in
Canada by J. B. Lippincott Company Ltd.,
Toronto

Printed and bound in Great Britain by
Aberdeen University Press

Contents

Preface	vii
1 The Importance of Diagnosis	1
2 The Nature of Disease and Diagnosis	9
3 The Issues of Reliability and Validity	27
4 Diagnosis as a Practical Decision-Making Process	49
5 Disease Entities in Psychiatry	60
6 International Differences in Diagnostic Criteria	70
7 The International Classification	86
8 The Role of Multivariate Analysis in Deriving or Validating Classifications	106
9 The Choice between Categories and Dimensions	119
10 Defining Diagnostic Criteria	137
11 Diagnosis by Computer	151
References	158
Index	167

Preface

Eighty years ago Hack Tuke observed that 'The wit of man has rarely been more exercised than in the attempt to classify the morbid mental phenomena covered by the term insanity', and went on to add that the result had been disappointing. The remark remains as apposite now as it was then, and goes far towards explaining why many present day psychiatrists have lost interest in the whole issue of diagnosis, while others have suggested that it is an unnecessary, even a harmful exercise. This book was born of the conviction that such attitudes are profoundly mistaken, and that the development of a reliable and valid classification of the phenomena of mental illness, and of the unambiguous diagnostic criteria which are essential to this task, are two of the most important problems facing contemporary psychiatry. Certainly the failure of both psychiatrists and psychologists to develop a satisfactory classification of their subject matter, or even to agree on the principles on which that classification should be based, is a most serious barrier to fruitful research into the aetiology of mental illness and even into the efficacy of therapeutic regimes. It is more exciting to develop explanatory theories, or to claim impressive results for this or that treatment, than it is to define the critical characteristics of the patients on whom one's research was based. It is probably more exciting to an architect to design parabolic canopies or baroque façades than it is to calculate the size and shape of the concrete slab on which his building will rest. But theories and therapeutic claims have no more chance of surviving than buildings if they are not built on secure foundations. Developing reliable diagnostic criteria and a valid classification may be as tedious as filling muddy holes with concrete but both provide the foundations on which all else depends.

It is not my purpose here to advocate any particular solution to the diagnostic controversies which have plagued psychiatry for so long, or to present a novel classification of my own. Part of the problem is that too many people have done this already. I have tried rather to discuss and clarify some of the conceptual and semantic problems involved, to explain why it has been so difficult to achieve adequate reliability, or even agreement on a common nomenclature, and to set

out the principles that should govern attempts to resolve these problems in the future. If we are to avoid repeating the mistakes of the past and a further cycle of failure and disappointment we must decide what we mean by terms like 'disease' and 'entity'; we must appreciate what diagnoses are, why they are necessary, and also what distortions they impose on our thinking; above all we must realize the importance of unambiguous operational or semantic definitions for all our diagnostic labels. We must also have some understanding of the perceptual and cognitive processes involved in the act of making a diagnosis; and if we are to make a rational choice between categorical and dimensional classifications, and appreciate the ways in which computers can and cannot help us, we must try to understand the various forms of multivariate analysis that have been applied to classificatory problems, and their many limitations.

Although this book is primarily an attempt to further these aims, it may also serve to illustrate the contemporary medical concept of mental illness to members of other professional groups. At all events, the assumptions made by behavioural scientists, and popular writers like Szasz and Laing, about what they call the 'medical model' have sometimes been inaccurate or outdated, and their arguments might carry greater conviction if they had a better understanding of the concepts they are criticizing, and the factual basis from which they are derived. Important as such misconceptions are, this book is aimed mainly at those directly concerned with the study and treatment of mental illness, and its central purpose is to remind them that diagnosis and classification are matters of fundamental importance, and to stimulate them to think about, and act on, the themes that are developed in the following pages.

Few books are ever written by their authors alone, and I cannot omit to record my gratitude to several of my colleagues in the preparation and revision of this one. In particular I would like to thank Sir Denis Hill, Professor A. E. Maxwell and Dr Norman Sartorius for their detailed and invaluable comments on the chapters in which I was really straying into the territory of their expertise, and Dr Norman Kreitman for his careful appraisal of the final manuscript. Above all, I am grateful to my secretary Liz Boxley for her efficiency, enthusiasm and patience.

R.E.K.
The Institute of Psychiatry
March 1974

1 The Importance of Diagnosis

In most branches of medicine the value of diagnosis is never questioned. Its importance is self-evident because treatment and prognosis are largely determined by it. If a man of forty with a cough and bloodstained sputum is diagnosed as having pulmonary tuberculosis, it follows almost automatically that he will be treated with isoniazid for several months, probably in combination with other anti-tuberculous drugs, and at the end of that time his health is likely to be restored. If, on the other hand, he is diagnosed as having a bronchial carcinoma his treatment and prognosis will both be quite different. He may have a pneumonectomy or radiotherapy, or neither, but either way he will not receive isoniazid. If he is fortunate his health may be restored, but it is much more likely that he will be dead within two years, and, depending on the site and extent of his disease, a fairly accurate estimate can be made of his chances of survival. It is also unlikely that there will be any disagreement about the diagnosis. The history and findings on physical examination may be equivocal, and the initial impressions of different doctors may conflict with one another. However, after a few days of investigations, chest X Ray, bronchoscopy, sputum examination and so on, the diagnosis will usually be beyond dispute, even if the patient has both illnesses.

Where mental illness is concerned the situation is rather different. A man of forty who is sleeping badly, cannot think clearly and suspects that his colleagues are talking about him behind his back may have either schizophrenia or a depressive illness. If he is schizophrenic, he is likely to be treated with phenothiazines, and to recover incompletely, or become chronically ill. If he has a depressive illness, he is more likely to be treated with ECT or a tricyclic anti-depressant, and to recover completely within two or three months. But whichever diagnosis he is given, he may still receive phenothiazines, or ECT, or a tricyclic drug, or even all three, and may recover completely, incompletely, or not at all. Psychiatrists are quite likely to disagree amongst themselves whether he has schizophrenia or an affective illness, and even about the definition and meaning of these two terms. They may even make a diagnosis of 'schizoaffective illness',

which to many clinicians must seem tantamount to making a diagnosis of 'tuberculoplasm' in a patient who has some symptoms suggesting that he has tuberculosis and others suggesting that he has a neoplasm.

This is the fundamental reason why the importance of diagnosis is not self-evident in psychiatry in the way that it is in other branches of medicine; because the therapeutic and prognostic implications of psychiatric diagnoses are relatively weak, and the diagnoses themselves relatively unreliable. Psychiatrists tend to react to this state of affairs in one of three ways. Those who are most strongly influenced by medical tradition usually minimize or deny the problem, insisting that diagnosis is crucial to rational treatment and sometimes striving to demonstrate clear relationships between diagnosis and therapeutic response. The majority admit the problem and react by taking less and less interest in the whole question of diagnosis. They continue, out of habit, to assign diagnoses to their patients, but the criteria they use become blunt and vague through neglect, and they are not disturbed when their colleagues' diagnoses differ from their own because they regard it as an 'academic matter' without any practical importance. As a result, diagnostic reliability falls further and a vicious circle develops as therapeutic and prognostic implications become even more tenuous and uncertain. A third group, people like Neumann in the last century and Karl Menninger in this, advocate abandoning diagnosis outright and argue persuasively the advantages, even the necessity, of doing so.

THE SHORTCOMINGS OF DIAGNOSES

There are indeed many problems and pitfalls associated with the act of assigning a diagnosis to a 'dis-eased' human being, particularly where psychiatry is concerned, and it may be as well to recognize these at the outset. In the first place, there are many situations in which a diagnosis seems almost pathetically inadequate to convey what we feel to be the essence of the patient's predicament. Karl Menninger, the most vocal and influential of contemporary abolitionists, has expressed the feeling very clearly (Menninger, 1948):

'What shall we call the "disease" represented by a man who has always been frail but has worked hard to support his widowed mother, did not feel he could afford to get married, buries himself in the details of a complicated job, develops paralysing headaches, loses time at the office for which pay is deducted from his wages, worries about this so much that he loses sleep and begins vomiting after each meal? Just to make it complicated he has a leukocytosis and an enlarged spleen. Does not such a disease defy diagnosis?'

'Even in the simplest cases it seems to me misleading to make a diagnosis in the old-fashioned way. A middle-aged puritan spinster appears in my office with a chancre on her lip. Isn't this a simple diagnosis? I don't think so. Nor would you if I told you the circumstances of how she acquired that chancre, whom she

acquired it from, how she happened to select that type of man, or why she permitted him to kiss her. Her sickness cannot be accurately diagnosed just as syphilis. She did not come to me because of it. What she came to me for was a more serious thing. She was so depressed about the implications of the infection that she now wanted to kill herself. What is the name of that disease? . . . What is the diagnosis in a patient who has coronary symptoms whenever he takes his wife to a party? Or in a woman who has migraine on the weekends that her son is home from college? What kind of arthritis is it which becomes activated with each quarterly meeting of the board of directors?'

Not only does a formal diagnosis often fail to convey what we feel to be the essential elements of the problem, it may not even tell us what the patient's symptoms are, or how he came to medical attention, or how he should be treated, or even what is likely to happen to him without treatment. To be told that a patient has an anxiety neurosis, for instance, does not tell us why he is frightened, or why he consulted a doctor, whether he is tense all the time or terrified episodically, whether drugs, interpretive psychotherapy or any of numerous forms of behavioural treatment are likely to help him, how badly disabled he is, or whether he is likely to recover without treatment. A related problem is that the majority of patients do not conform to the tidy stereotyped descriptions found in textbooks. They possess some, but not all, of the symptoms of two or three different diagnostic categories and so have to be allocated more or less arbitrarily to whichever syndrome they most nearly resemble. As a result disagreements about diagnosis are frequent and patients' diagnoses change repeatedly as they move from one doctor or hospital to another.

THE HARMFUL EFFECTS OF DIAGNOSES

There are other equally important, if less tangible, objections to diagnoses, as Menninger has emphasized. Attaching a name to a condition creates a spurious impression of understanding so that we cease to be puzzled and to ask questions. To say that someone has schizophrenia really says little more than that he is behaving in a rather odd way which we have encountered in other people in the past, but to the layman, and to some doctors also, it implies that we understand what is wrong with him, that he has an 'illness' like measles or appendicitis whose cause we either know already or will soon discover, and that there is a fundamental difference between this and other kinds of odd behaviour. Hardin (1956) coined the word *panchreston* (meaning 'explain-all', by analogy with panacea, or 'cure-all') to draw attention to the ways in which we use our jargon to provide comforting but meaningless explanations for things we really do not understand. He did us a valuable service by doing so, but there may be other dimensions to the problem as well. It is worth recalling that the belief that names are invested with magical powers was once widespread amongst our ancestors. There are

many surviving accounts and legends of kings and gods having special names known only to themselves and their most devoted followers and, as the stories of Aladdin and Rumpelstiltskin illustrate, it was believed that anyone discovering the secret name thereby achieved power over its owner. It is perhaps not too fanciful to suggest that the lingering shadow of this belief is part of the reason why doctors and patients alike find it so comforting to have impressive Greek or Latin names for the sicknesses that oppress them. Doubtless there are also more mundane reasons why doctors choose to conceal their ignorance from their patients and their relatives by clothing it in a Greek neologism, but all too often they also deceive themselves and treat the disease instead of the patient.

This last effect, the subtle way in which making a diagnosis distracts the doctor from his primary function of trying to relieve the suffering of a 'dis-eased' person and encourages him to treat an inanimate disease instead, is the most important drawback of all in the eyes of many psychiatrists. They feel, with some justification, that, whatever other branches of medicine may do, psychiatry must be concerned with the patient as a person, including his hopes and fears, his memories and daydreams, and that to attribute a diagnosis to him is inevitably to dehumanize him and to deflect attention from him onto his biochemistry or his genes or his delusions. Like Adolph Meyer (1907), they insist that psychiatrists should be concerned with understanding the sick person in terms of his life experience rather than with fitting his symptoms into a classificatory scheme. Others go further, maintaining that it is demeaning to any human being to be labelled like a specimen in a museum. Although it may be a little fanciful to claim that all labelling is necessarily derogatory it is certainly true, as Albee (1970) and others have pointed out, that the act of assigning a diagnosis to a patient inevitably focuses attention on his deficiencies rather than on his assets, and also that many psychiatric diagnoses have strong pejorative connotations. It is by no means rare for the aura surrounding such terms as hysterical, neurotic, psychopath and schizophrenic to have harmful effects on people's behaviour and attitudes towards the patient so labelled, and on his own attitude to himself; and at times they are used, by professional personnel as well as by laymen, as little more than thinly disguised expressions of contempt. What is more, labels like schizophrenia are sometimes attached to people on wholly inadequate evidence, and once attached may be almost impossible to remove (Rosenhan, 1973).

THE CALL FOR THE ABANDONMENT OF DIAGNOSIS

In the face of this impressive catalogue of shortcomings, misuses and pitfalls it is hardly surprising that some critics have advocated doing away with diagnoses completely. In Menninger's words: 'We affirm the necessity of cutting the Gordian Knot and using no names at all for these conditions of mental illness' (Menninger, 1963). Instead, he and other psychoanalysts would substitute a

lengthy formulation of each patient's predicament, describing his symptoms and their evolution, the strengths and weaknesses of his character, the nature of his relationships with other people, the stresses he was under and the way in which these played on his weaknesses and reopened old wounds, and the way in which his current behaviour and attitudes represent an attempt to defend himself against internal stresses. Every such formulation would be different, just as every human being is different, and would incorporate a plan of treatment tailored to the needs of that unique individual. (In fact, Menninger and many other psychiatrists still refer to this as a diagnosis; the term formulation is used here instead to avoid the ambiguity which would be bound to arise if the same word was used for the traditional label and for this comprehensive assessment.)

Essentially, Menninger's argument is that, because in one sense all mental illness rests on a single common foundation shared by all mankind, while in another equally important sense every individual is unique, and unique in his sickness, it is useless to classify either patients or illnesses; the sick individual must be assessed and treated on his own merits free from the pernicious restraints imposed by disease categories.

The fallacy in the argument

This argument is based on a serious fallacy, which soon becomes apparent if we consider the functions of classification and class membership. Whatever the context, whether one is concerned with sickness or not, there are three aspects to every human being:

1. those he shares with all mankind
2. those he shares with some other men, but not all
3. those which are unique to him.

The value of classification in any given context depends on the size of the second of these categories relative to the other two. The larger the first category the less the need for classification, and the larger the third the less the value of any classification that is attempted. Implicit in Menninger's viewpoint is the assumption that where mental illness is concerned, the first and third of these categories overshadow the second, and that the second can therefore be ignored.

Let us take this assumption at face value to see where it leads. In so far as all men are the same we cannot distinguish between one type of mentally ill person and another. Indeed, we cannot even distinguish between those who are ill and those who are healthy. It follows from this that if we have more than one form of treatment at our disposal we can have no rational criteria for employing treatment A in one situation and treatment B in another. We cannot even discriminate between those who need treatment and those who do not. On the other hand, in so far as every individual is unique, all learning from experience and all useful communication with others are rendered impossible. If every patient

is different from every other then we can learn nothing from our colleagues, our textbooks, or the accumulated experience of our predecessors. We cannot even learn from our own personal experience if there are no significant similarities between our last patient and the next. We may increase in skill and understanding in the course of treating one individual, but that skill and understanding will not carry over to the next. In short, insistence on a unitary concept of mental illness condemns us to giving the same treatment to every one, and prevents us even distinguishing between sickness and health. Insistence on the uniqueness of every individual prevents all learning and all communication concerning disease.

In fact, it is impossible to avoid classifying patients unless one is prepared to accept the facile Rogerian dictum that 'Therapy is good for people. Period.' and to offer the same panacea to everyone. As soon as one begins to recognize features that are common to some patients but not to all, and to distinguish those which are important from those, like eye colour, which are not, one is classifying them, whether one recognizes it or not. The only point at issue is what sort of classification one is going to have. As Pasamanick (1963) has pointed out, even our language is based on classification; every common noun expresses the recognition of a class. Menninger may well be correct in maintaining that, in comparison with physical illness, category (2) above is small where mental illness is concerned relative to categories (1) and (3), but we still have to focus our attention on category (2) if we are ever to acquire any useful understanding and pass that understanding on to others.

THE UNAVOIDABILITY OF DIAGNOSIS

In fact, the detailed personal formulation advocated as an alternative to a diagnosis is no alternative at all. The two are required for quite different purposes, and have different functions. A formulation which takes into account the unique features of the patient and his environment and the interaction between the two is often essential for any real understanding of his predicament, and for planning effective treatment, but it is unusable in any situation in which populations or groups of patients need to be considered. In any such situation some form of classification or categorization is unavoidable. It is important not to see this issue simply as a controversy between psychoanalysts and 'organic' psychiatrists. Feinstein and Kline, starting from quite different theoretical positions from Menninger, have also commented forcibly on the shortcomings of diagnoses in isolation, and argued convincingly that they need to be augmented by other data for research purposes just as much as in the management of the individual patient (Kline, Tenney, Nicolaou and Malzberg, 1953; Feinstein, 1964). But it is one thing to supplement a diagnosis with other information and quite another to abandon it. Without diagnosis, or some comparable method of classification, epidemiological research would be impossible. We would have no way of finding

out whether mental illness was commoner in one culture than another, or whether its incidence and manifestations varied with other factors like poverty, social class and ethnic background. Without a criterion for distinguishing between sickness and health, and between one sort of sickness and another, there could never be any rational planning of psychiatric services. Indeed all scientific communication would be impossible and our professional journals would be restricted to individual case reports, anecdotes and statements of opinion.

The clinician, who deals with one patient at a time, may succeed in convincing himself that he does not and need not classify his patients, but even here, as we have seen, this is not really so unless he is offering the same treatment to everyone he sees. The research worker and administrator, who deal with populations of patients rather than individuals, cannot even pretend to avoid classification. There is no point in defining a population unless its members possess something in common with one another which distinguish them from members of other populations, and once this condition has been satisfied a classification has been created.

I would like to emphasize, though, that this argument, which I believe to be irrefutable, is in no sense a denial of the value of the comprehensive formulation of the individual patient's predicament which Menninger and others have advocated, nor is it a brusque dismissal of their criticisms of the shortcomings of our existing classifications and of the ways in which we use them. It is undeniable that a diagnosis *by itself* is almost never an adequate basis for treating an individual patient. It will probably set limits to what it is possible, or necessary, to achieve, and exclude some forms of treatment from consideration, but a host of other factors, the patient's personality, his relationships with other people, his reasons for seeking treatment, his social background and so on, will all influence one's aims and the ways in which one seeks to achieve them. The complexity of the situation can only be dealt with adequately by a comprehensive analysis.

SHORTCOMINGS AND HARMFUL EFFECTS IN PERSPECTIVE

However, not all the criticisms that have been levelled at psychiatric diagnoses are valid, and others have been overstated. It is true that a diagnosis may provide little definite information about the patient's prognosis' or even about his symptoms or how he should be treated, but at the very least it has the important negative function of excluding from consideration many other types of problem. To know that a patient has a depressive neurosis at least tells us that he is not elated, or deluded, or hallucinated; that the question of compulsory detention in hospital does not arise; and that he is sad and unhappy for reasons that seem understandable. It is undeniable that psychiatric diagnoses are often unreliable, and that it is commonplace for a single patient to be given three different diagnoses by three different psychiatrists, but this does not have to be so. It has been

demonstrated several times that adequately trained psychiatrists can achieve acceptable levels of agreement, and that there are consistent differences in symptomatology, course and response to treatment between populations of patients from different diagnostic categories. It has to be admitted that the names we give our illnesses sometimes encourage facile assumptions about disease entities and causes, what Cohen (1953) called a 'penny in the slot machine' approach to diagnosis, but this problem also can be dealt with by educating doctors without abandoning diagnosis. As to the belief that attributing a diagnosis to a patient somehow detracts from his dignity as a human being, it is surely the role that matters rather than the diagnosis. It is being a patient, and particularly a psychiatric patient, that hurts, rather than being regarded as a manic depressive. Without changing society's attitudes to mental illness and other forms of deviance it is difficult to see how abandoning diagnoses would help very much. Although it is true that several of our diagnostic terms, like hysterical and psychopath, have acquired pejorative connotations even amongst psychiatrists, these connotations exist because they denote ways of being mentally ill, or failing to cope. It is the fact that the patient is crazy, or manipulative, or unable to cope with the demands of everyday life, that creates the odium, not the diagnosis itself. And, although the use of diagnostic labels may indeed make it easier for psychiatrists to lose sight of the fact that their patients are also people with ordinary human feelings, they also enable them to recognize them as schizophrenics or manic depressives and give them the appropriate treatment. Acceptance and understanding are valuable, but so too is recognition and if there has to be a choice between the two, the latter may well be more important in the main areas of psychiatric practice.

CONCLUSIONS

In summary then, psychiatric diagnoses have serious shortcomings. They sometimes bring other disadvantages in their train and they are liable to be misused in various ways. But none of these is an intrinsic defect of all classifications, nor is any tool safe from the risk of misuse. As with any other innovation, the advantages and disadvantages of a classification of psychiatric disorders have to be balanced against one another. The advantage of a classification is quite simply that it allows us to communicate.

The following chapters take these conclusions as their starting point. From this point on it will be taken as proven that psychiatry cannot function at all without classifying its subject matter. It will also be regarded as self-evident that our present classifications are imperfect, and also misused. What follows is essentially a description of our existing Kraepelinian classification, with its strengths and weaknesses, its uses and failings, followed by a discussion of some of the ways in which it could be improved, and an outline of some of the alternative forms of classification with which it might be replaced.

2 The Nature of Disease and Diagnosis

DEFINITIONS OF DISEASE

There is no concept in medicine more fundamental than that of disease or illness. Everyone, physician and layman alike, uses the words and takes it for granted that their meaning is self-evident and unambiguous. Perhaps because they are so commonplace little thought is ever given to them and they are rarely defined except in dictionaries. Occasionally, arguments develop about whether or not a particular phenomenon, like alcoholism or homosexuality, is a disease, but even these rarely lead to a search for a definition. Laymen tend to feel uneasily that only physicians can be expected to answer such questions; and physicians are either so convinced that the answer is yes, or alternatively no, that no thought is necessary, or else plead that no answer is possible in our present state of knowledge. Rarely does anyone suggest, particularly to a medical audience, that the answer depends on what is meant by disease. In fact, as Scadding (1963) has pointed out, all sorts of problems result from our failure to define fundamental terms clearly enough, and the worst confusion is always caused by terms that are so familiar that we are hardly aware that any assumptions are involved in using them.

Even when allowance is made for the essentially empirical nature of medicine, and the impatience with theorizing and philosophizing that such an atmosphere engenders, it is remarkable how little thought has been given to deciding exactly what 'disease' is, and what properties diseases should and should not possess. Most physicians rarely give such matters a moment's thought, even though they may well unwittingly use the words disease and illness in different senses at different times, or get involved in disputes produced by unrecognized differences in the meaning attributed to them. Psychiatrists are usually more aware than others of some of the problems involved, because the forensic decision to remove the burden of criminal responsibility from the mentally ill frequently involves them in having to decide whether people are ill or not, and to justify their decisions to hard-headed lawyers; but they have not been conspicuously successful in finding solutions. It is important to appreciate, though, that these problems do not apply only to psychiatry. The whole of medicine is in much the same predicament, or would be if important legal decisions hinged on the issue.

It is not so much a definition we lack as an adequate one. Numerous definitions of disease have been suggested from time to time, but the majority are either hopelessly vague, or tautologous, or exclude people who, by common consent, are ill or include others who, by the same token, are not. Most of the definitions offered in dictionaries, including medical ones, define disease either as a state of sickness or impaired health, which hardly advances the matter very far, or in terms of a disturbance of function, leaving it unclear how the presence of a disturbance is to be recognized. The definition in the current (24th) edition of Dorland's *Illustrated Medical Dictionary* ('A definite morbid process having a characteristic train of symptoms . . .') is representative of the former, and that in the *Shorter Oxford English Dictionary* ('a condition of the body, or of some part or organ of the body, in which its functions are disturbed or deranged') of the latter.

DISEASE AS SUFFERING

Historically there can be little doubt that the concept of disease originated as an explanation for the onset of suffering and incapacity in the absence of obvious injury, that the concept of health was a later development, implying the absence of disease. Naturally enough, therefore, attempts have often been made to define illness in terms of suffering or incapacity, or at least in terms of a complaint of some sort. However, this immediately leads to difficulties. Many people who are ill, even physically ill, do not complain or suffer, either because they experience no symptoms, or because they ignore what in others would be cause for complaint, or simply because they drop dead without warning. A man with a cancer growing silently in his lung, or someone with cardiac pain which he dismisses as 'a touch of wind' would both be regarded by doctors and laymen as ill, and urgently in need of treatment, yet neither complains, or even suffers to any significant extent. The same is true of the manic patient who has never felt better in his life, or the schizophrenic who is unshakably convinced that his voices are real, or the typhoid carrier harbouring salmonellae in his gall bladder. Other people complain incessantly, and insist that they suffer, without either their doctors or anyone else being convinced that they are genuinely ill.

DISEASE IS WHAT DOCTORS TREAT

Partly because of problems of this sort Kraupl Taylor (1971) has suggested that what he calls 'therapeutic concern' would be a more appropriate criterion; people would be regarded as ill, in other words, if either they or others were convinced they needed medical attention. A criterion of this sort would certainly be capable of embracing people whom doctors, or society as a whole, regarded as in need of treatment as well as those who complained or suffered personally, but in doing so it would create a worse problem than the one it solved. Equating illness with a complaint allows the individual to be sole arbiter of whether he is

ill or not. This is unsatisfactory because some people who should be complaining don't do so, and others who do so repeatedly don't seem to have adequate reasons for doing so. Equating illness with other people's 'therapeutic concern' implies that no one can be ill until he has been recognized as such, and also gives doctors, and society, free rein to label all deviants as ill, thus opening the door to all the inconsistencies and abuses Szasz has vividly conjured up.

The fact is that any definition of disease that boils down to 'what people complain of' or 'what doctors treat', or some combination of the two, is almost worse than no definition at all. It is free to expand or contract with changes in social attitudes and therapeutic optimism and is at the mercy of idiosyncratic judgements by doctors or patients. If one wished to compare the incidence of disease in two different cultures, or in a single population at two different times, whose criteria of suffering or therapeutic concern would one use? And if the results of a study comparing the amount of sickness in a society at two different times suggested that the incidence of illness had risen, would this be because people's health had deteriorated, or because their attitudes to illness had changed?

DISEASE AS A LESION

The development first of morbid anatomy and then of histology in the 19th century produced widespread evidence that illness was often accompanied by structural damage to the body, either at a gross or a microscopic level. It was only a short step from this observation to the assumption that these lesions constituted the illness, and that illness necessarily involved structural damage. Subsequently, as knowledge of biochemistry and physiology caught up with that of anatomy, this concept was expanded to include evidence of a biochemical or physiological abnormality, without relinquishing the basic assumption that illness necessarily involved a demonstrable physical abnormality of some sort.

In this 19th century milieu it was almost inevitable that the presence of an identifiable lesion should come to be regarded as the hallmark or defining characteristic of disease, and this concept of illness was, explicitly or implicitly, dominant until quite recently. Such a standpoint does indeed have many advantages. It provides an objective criterion which is not at the mercy of changing social attitudes and therapeutic fashions, and also embodies at least a partial explanation of the patient's symptoms or disabilities. On close examination, however, a number of problems arise. The concept of an abnormality or a lesion is quite straightforward so long as one is concerned with a departure from a standard pattern. But as soon as we begin to recognize that there is no single standard pattern of structure or function, and that even healthy human beings and all their constituent tissues and organs vary considerably in size, shape, composition and functional efficiency, it becomes much less obvious what constitutes a lesion, where normal variation ends and abnormality begins. Is, for

instance, hypertension a disease, and if so what is the level beyond which the blood pressure is 'abnormal'? Is shortsightedness to be regarded as a disease? And what of minor congenital abnormalities, like fused second and third toes and albinism? An even more serious problem is that symptoms whose physical basis has not yet been demonstrated cannot be accepted as diseases. Trigeminal neuralgia, senile pruritus, proctalgia fugax, dystonia musculorum, even migraine, must all be discarded. Fifty years ago, the same would have been true of most forms of epilepsy, Parkinson's disease, chorea, Bornholm disease and most deliria.

Insistence on the presence of a demonstrable lesion causes particular problems for psychiatry as no physical basis has yet been identified for most of its major syndromes. For this reason, the majority of psychiatrists have balked at accepting any such criterion, though a few have done so with enthusiasm. Kurt Schneider (1950) willingly accepted that the word illness should only be used in situations in which 'some actual morbid change' or 'defective structure' was present in the body, and stated bluntly that he did not regard either neurotic states or personality disorders as illness, but simply as 'abnormal varieties of sane mental life'. However, he still regarded schizophrenia and manic depressive psychosis as illness, along with organic and toxic psychoses, on the basis of a convenient but unjustifiable assumption that in time they would prove to have an 'underlying morbid physical condition'. Szasz (1960) has gone further, taking Schneider's argument to its logical conclusion and maintaining that, as they lack any demonstrable physical basis, there cannot be any such thing as mental illnesses and that to speak of illness in this context is to use the word in a purely metaphorical sense.

DISEASE AS ADAPTATION TO STRESS

For understandable reasons, the reaction against the 19th century's 'physical lesion' concept of illness was led by psychiatrists, particularly by Adolph Meyer in Baltimore. Meyer insisted that diseases, especially mental ones, were reactions of the whole organism to its total environment, rather than noxae attacking it from without or structural lesions within. He was also at pains to emphasize that, as every individual is unique, and his environment likewise, the reaction which constituted his disease must also be unique. The theme was taken up enthusiastically by his numerous pupils. Seguin (1946) proposed that disease should be formally defined as 'a reaction of the organism as a whole to external or internal stimuli altering seriously its equilibrium' and reiterated that a patient's disease 'can only be considered as a unique phenomenon, and handled as such'.

A definition of this sort, emphasizing the differences between one sick individual and another rather than their similarities, obviously hampers attempts

to classify diseases, but an even more serious problem is that it fails to provide any criterion for distinguishing between sickness and health, or disease and non-disease. It would be quite reasonable to define life itself as a series of reactions by the organism to external or internal stimuli. How then is one to distinguish between those which 'alter seriously its equilibrium' and those which do not? Seguin's own criterion that equilibrium is seriously altered 'when the organism cannot recover from it' is clearly inadequate, as it would appear to exclude all illnesses normally ending in full recovery. Alternatively, if it does not do this, it provides no means of distinguishing between temporary disturbances which are to be regarded as disease and those, like the breathlessness, tachycardia and metabolic acidosis produced by exercise, which are not.

DISEASE AS IMPERFECTION

Health and sickness, like good and evil, and light and dark, constitute a single bipolar concept in the sense that they are the antithesis of one another. The presence of either excludes the other or, more precisely, any increase in the one diminishes the other.

This means that whenever we determine the extent or degree to which someone is sick, in doing so, we automatically determine how healthy he is as well. It also means that concepts or definitions of sickness carry with them implicit concepts or definitions of health, and vice versa. If sickness is defined by the presence of a 'lesion', any human condition in which no lesion is to be found is by definition healthy. And if all departures from normality, all imperfections, are regarded as forms of illness, then the only truly healthy state is one of perfection. The most influential conceptual system of this latter kind is psychoanalytic theory. According to this theory, the unavoidable vicissitudes of early life activate morbid psychological mechanisms which impair personality development and functioning in everyone to a greater or lesser extent. Although these handicaps are regarded as potentially reversible if appropriate treatment is provided, this implicitly reduces health to an idealized state of complete normality, an almost hypothetical state to which man can only aspire, and some degree of sickness becomes his universal lot. There is, to be sure, nothing inherently absurd or illogical in this, but it does carry profound implications for society's attitudes towards the sick, for individuals' attitudes to themselves, and not least for the staffing and financing of medical services. On all these counts a more restricted concept of sickness seems preferable. The definition of health as 'a state of complete physical, mental and social wellbeing' enshrined in the charter of the World Health Organization has similar implications, although in view of its setting, it is probably wiser to regard this particular statement as a fine ringing phrase intended to epitomize the political aspirations of a great international organization rather than as a practical definition.

A STATISTICAL CONCEPT OF DISEASE

Although psychiatrists were the first to protest about the short-comings of the 19th century's lesion/pathogen concept of disease, these eventually became apparent to others also. When Pickering and his colleagues (Oldham, Pickering, Fraser Roberts and Sowry, 1960) demonstrated beyond reasonable doubt that such a major cause of death and disability as essential hypertension was a graded characteristic, dependent like height and intelligence on polygenic inheritance and shading insensibly into normality, it was clear that a new concept of illness was required, and that this would have to be based on a statistical model of the relationship between normality and abnormality. In fact, Cohen (1943) had foreseen this some years before by defining disease simply as 'deviation from the normal . . . by way of excess or defect'. Indeed, Broussais and Magendie had come close to doing the same thing a hundred years earlier. But in spite of its statistical basis – from the general tenor of his essay – one must assume that he was using the term normal in a statistical rather than in an ideal sense – Cohen's definition is still inadequate. This is mainly because, as it stands, it does not distinguish between deviations from the norm which are harmful, like hypertension, those which are neutral, like great height, and those which are positively beneficial, like superior intelligence. Scadding (1967) has attempted to deal with this problem by stipulating that the abnormal phenomena must place the affected individual at a 'biological disadvantage', reducing either his chances of survival or his chances of procreation. As the formal definition of disease he offers is based on a very careful consideration of the many problems and issues involved, and is certainly less unsatisfactory than any of the alternatives mentioned previously, it is worth studying in detail.

SCADDING'S DEFINITION

Scadding defines a disease as 'the sum of the abnormal phenomena displayed by a group of living organisms in association with a specified common characteristic or set of characteristics by which they differ from the norm for their species in such a way as to place them at a biological disadvantage'.

Although this is a definition of a disease, rather than of the global concept of disease, it has a number of fundamental implications for the latter. The most important of these is that it is based on Cohen's idea of deviation from the norm of the species, set out in more explicitly statistical terms. This statistical basis is the crux of the matter, for it carries with it several fundamental implications – that deviation in either direction, too much or too little, is equally capable of producing disease; that the boundary between sickness and health may need to be an arbitrary one, like the boundary between mental subnormality and normal intelligence; and that the majority are automatically debarred from being re-

garded as ill. The 'specified common characteristic or set of characteristics' is the defining characteristic of the individual disease in question. Its presence is essential for establishing the presence of that disease, and it is worth noting in passing that Scadding's wording allows it to be either monothetic (a single trait) or polythetic (a set of traits, no one of which is mandatory). The concept of biological disadvantage is obviously a useful one, though it is less unambiguous than it seems. Although any increase in mortality or reduction in fertility are presumably adequate to establish its presence, it is unclear whether disabilities which are either too slight to affect mortality or too late developing to impair fertility can also qualify. Moreover, what is a disadvantage to the individual may not always be so to the population to which he belongs, and it is not clear whose disadvantage, the individual's or the group's, should take precedence. Problems of this sort are particularly important where social animals like man are concerned. Homosexuality, for instance, is undoubtedly a biological disadvantage to the individual by virtue of drastically reducing his chances of procreating. But if it could be shown that homosexuals possessed valuable aptitudes which others lacked, it could be argued quite plausibly that a community which contained a proportion of homosexuals might have competitive advantages over those that did not. Indeed, in an era of explosive population growth like that which our species is currently experiencing, it may be biologically advantageous to a community to have its fertility reduced anyway.

This is simply one example of a more general problem. A characteristic which is a disadvantage in one environment may be beneficial in another. The sickle cell trait is a deviation from the norm which in most environments produces a slight but definite biological disadvantage. In an environment in which malaria is endemic, however, it is positively beneficial. Is it then to be regarded as an illness in the first environment but not in the second? Such problems as this arise because the environment, with its capacity to tilt the scales one way or the other, is ignored in Scadding's definition. This is a particularly serious matter where mental illness is concerned because here the environment, and especially its social aspects, is often of paramount importance. Qualities like recklessness and aggressiveness, for example, may lead a man to be regarded as a psychopath in one environment and to be admired as a hero in another. Of course, it could be argued very reasonably that social advantages or disadvantages of this kind are irrelevant, and that the disadvantage is specified as being a biological one for this reason. But it is not always easy to decide where a biological disadvantage ends and a social one begins, or to visualize a human environment shorn of all social components.

The key phrase in Scadding's definition, 'the sum of the abnormal phenomena', may also be difficult to apply where mental illness is concerned. For most physical illnesses, the norm is fairly clearly established because the relevant physiological mechanisms are reasonably well understood. We know a great deal

about the normal functioning of the heart and kidney and so find it relatively easy to decide what is abnormal and what is not. We know much less about normal psychological functioning and find it correspondingly harder to distinguish between the normal and the pathological. It is more difficult than we like to admit, for instance, to distinguish between normal grief and a pathological reaction to bereavement, or between a delusion and an overvalued idea. Where an abnormality of part function is involved, as in these two examples, the situation is not too desperate. We have psychological criteria of a sort available for assessing normality. However, where the only abnormal phenomena involve the total behaviour of the individual, and the abnormality is essentially a departure from a social or moral norm, we are soon adrift. As Lewis (1953) has emphasized, once we allow social norms to become our criterion of normality all socially deviant or disapproved behaviour becomes liable to be regarded as sickness. Lewis's own solution to this quandary was to insist that our criterion of abnormality where mental illness was concerned should be firmly rooted in a disturbance of psychological functioning, in the same way that our criterion of physical abnormality is rooted in a disturbance of physiological function. The snag is that at present we know so little psychology that this course can be little more than an aspiration. In many areas, if we are determined not to resort to social criteria, all we can do is defer judgement.

Many of these problems are obviously caused by the short-comings of psychology and psychiatry, rather than by those of Scadding's definition. But they do illustrate the difficulties of applying even so carefully thought out a definition as this within the realm of mental illness. Certainly, they provide little grounds for confidence in expecting formal definitions of disease to provide us with a firm criterion for deciding whether the psychopath and the exhibitionist are ill or not, or for distinguishing the phobias, mood changes and rituals of neurotic illness from the fears, fancies and idiosyncrasies of normal health.

DISEASE AS A PLAN OF ACTION

An alternative way of elucidating the meaning of the word disease, or illness, is to examine how it is used in practice. If one does this, it soon becomes apparent that there are many situations in which everyone is agreed that a disease is present, and others where the question of illness does not arise at all. But there are also a number of areas in which physicians disagree amongst themselves. Where mental illness is concerned, the main disputed area is the general field of personality disorder, together with the related phenomena of sexual deviation, alcoholism and drug abuse. Physical disorders generally give rise to less disagreement, but a symptom-free typhoid carrier, a woman with a carcinoma-in-situ of the cervix, a child with a congenital club foot, or anyone complaining of pain for which no cause can be found, are all liable to be regarded as ill by

some doctors but not by others. In such cases, it will almost invariably be found that those who regard the subject as ill also regard some medical procedure, treatment or investigation, as necessary, while those who do not regard the subject as ill do not regard either treatment or investigation as warranted. This gives rise to the suspicion that the answer one gets to the question 'Is this a disease?' is really a covert answer to the quite different question 'Should this person be under medical care?' As Linder has put it, a diagnosis is not a description, an explanation, or even a prediction, it is simply a disguised prescription (Linder, 1965).

Linder has also drawn attention to one significant exception to this principle which is of particular importance to psychiatry. In any situation in which a man's behaviour infringes the law, or is offensive to others in more general ways, he is normally liable to be blamed, or even punished, but if he is deemed to be mentally ill this is no longer so. Explicitly in the case of the criminal law, and in practice also in the case of public attitudes, mental illness excuses behaviour which would otherwise invite blame and retribution. In any setting in which this issue arises, in a court of law or not, physicians may disagree about whether the subject is ill because they disagree about whether it is appropriate for him to be held responsible for his behaviour. In other words, there are some settings in which the statement that an individual is or is not ill may be a covert answer to the question 'Should he be held responsible for his behaviour?' rather than to the usual question 'Should he be under medical care?'.

Although this formulation takes no account of the routine involvement of physicians in essentially physiological processes like childbirth, and more recent incursions into other areas like family planning, it is difficult to refute Linder's basic argument. Indeed, in some ways it clarifies the problem by implying that disagreements about what is or is not illness are not really logical or semantic disagreements at all, but simply empirical disagreements over the issue of whether, in a given case, medical procedures are likely to be beneficial. Among other things, it explains why such disagreements are commoner in psychiatry than in other branches of medicine. They are more common simply because there is more uncertainty about the efficacy of the therapeutic measures employed in psychiatry than there is elsewhere. If, for example, the effectiveness of the various techniques used in the treatment of personality disorders was established one way or the other, we would probably no longer disagree about whether personality disorders were illnesses or not.

This view of diagnosis as a disguised plan of action obviously has much in common with the 'disease is what doctors treat' argument, or Kraupl Taylor's criterion of 'therapeutic concern', but there is a crucial difference. Kraupl Taylor is advocating the presence or absence of 'therapeutic concern' as a workable criterion for the presence or absence of illness or, as he would call it, 'patienthood'. Linder is arguing that we have no consistent criterion for the

presence of illness, but disguise the fact by substituting the purely practical issue of whether or not medical attention is appropriate.

Scadding's definition or Linder's argument?

Assuming that this is an accurate analysis of the situation, we are faced with a choice between adopting Scadding's definition of disease, or some modification of it, and trying to develop a set of working rules from this statement of principles, or alternatively accepting Linder's argument that we do not possess any concept of illness other than the purely empirical concept of what, in any given state of therapeutic optimism and ability, is a fit subject for medical attention. Which of the two we accept will almost inevitably be determined by the context. In any situation in which we need to measure the amount of illness in a population, either to compare one population with another or to measure change within a single population over time, we must have a definition of disease which is independent of therapeutic innovations and fashions. This means that we must use Scadding's definition, or something comparable, or abandon our task. But in the everyday situation where we are concerned with the subject matter of medicine, rather than with disease, we can afford to accept Linder's argument. This is so because in practice, it is not necessary to decide precisely what illness is in order to classify illnesses. This is well illustrated by the fact that none of the three glossaries to the current (8th) edition of the Mental Disorders section of the *International Classification*, American, British or WHO, makes any attempts to define 'mental disorder' in spite of the fact that they are all concerned with the classification of its various manifestations. Nor do they even remark on their failure to do so, though the British glossary does comment on its failure to define the term 'psychosis', adding significantly that 'no such definition is required for the effective use of the classification' (General Register Office, 1968). This singular omission might well strike taxonomists as surprising, even shocking, and certainly the lack of a definition allows the territory embraced by the terms mental illness and mental disorder to expand or contract in unpredictable ways. But it can legitimately be argued that, in practice, medical classifications need this freedom because they need to embrace all those whom physicians encounter in the patient role, regardless of whether they are ill or not. If otherwise healthy people whose marriages are threatening to break up frequently seek help from psychiatrists, then psychiatric classifications need to be constructed to accommodate such people, and the same applies to adolescents referred to psychiatrists by courts because they are taking drugs, or children brought by their parents for starting fires. The corollary to this, which is not as widely recognized as it should be, is that it does not necessarily follow because someone has been examined by a psychiatrist, allocated to a diagnostic category and offered treatment, that they are or ever were mentally ill.

DEFINITIONS OF INDIVIDUAL DISEASES

If we have no adequate definition of disease as a global concept, it becomes even more important to have adequate definitions for individual diseases. Unfortunately, as soon as one begins to consider individual illnesses, it becomes apparent that there is no consistent theme. Some, like tuberculosis, are defined by their cause, others, like ulcerative colitis, by their pathology, others, like migraine, by their symptoms, and so on. The reasons for this state of affairs lie in the historical development of the global concept of disease we have just been considering. To the Cnidean School of the ancient world symptoms and signs were themselves diseases. Fever, asthma, joint pains and skin rashes were all separate diseases to be studied individually, and this assumption persisted until very recently. Most of the 2400 diseases Boissier de Sauvages described in the 18th century were merely individual symptoms. The idea of disease as a syndrome, a constellation of related symptoms with a characteristic prognosis – to remit, to evolve or to persist – originated with Sydenham in the 17th century, though the Greek Empiricist school had had the germ of the idea long before. This concept lasted until the early 19th century when Morgagni and Bichat popularized post mortem dissection of the body as a routine procedure, and so converted disease from a clinical entity observed at the bedside to a characteristic morbid anatomy observed in the cadaver. Thereafter, new concepts followed one another in rapid progression mainly, as Feinstein (1969) has pointed out, in response to the introduction of new types of observational technology. With the development of powerful microscopes in the middle of the 19th century, individual cells could be examined as well as tissues and whole organs, and the consequent detection of cellular pathology led Virchow and his contemporaries to assume that cellular derangements were the basis of all disease. This concept was, in its turn, displaced by the discovery of bacteria by Koch and Pasteur, a development that was responsible more than anything else for the concept of 'disease entities' each produced by a single aetiological agent. Currently, new techniques like electrophoresis, chromosomal analysis and electron microscopy are producing yet further concepts of disease expressed in terms of deranged biophysical structures, genes and molecules. As Riese (1953) aptly observed, the history of medicine could be written in terms of man's changing concept of disease.

Each of these waves of technology has added new diseases and from each stage some have survived, so the diseases which figure in contemporary textbooks have a very variable conceptual basis. A few, like senile pruritus and proctalgia fugax, are simply Cnidean symptoms, even though first described relatively recently. Others, like migraine and most of the diseases of psychiatry, are clinical syndromes, Sydenham's constellation of symptoms. Mitral stenosis and cholelithiasis are based on morbid anatomy, and tumours of all kinds on histopathology.

Tuberculosis and syphilis are based on bacteriology and the concept of the aetiological agent, prophyria on biochemistry, myasthenia gravis on physiological dysfunction, Down's syndrome on chromosomal architecture, and so on. In fact, our present classification is rather like an old mansion which has been refurnished many times, but always without clearing out the old furniture first, so that amongst the new inflatable plastic settees and glass coffee tables are still scattered a few old Tudor stools, Jacobean dressers and Regency commodes, and a great deal of Victoriana. Indeed, Scadding is probably close to the truth when he suggests that it is this logical heterogeneity in our definitions of individual diseases that is responsible for our inability to produce a satisfactory definition of disease as a whole (Scadding, 1972).

An even older and more influential relic from the past is the assumption that diseases are intangible noxae, which silently and mysteriously attack the individual from without. All cultures recognize at least some forms of incapacity and suffering as disease and most, including our own until recently, attribute these to malign supernatural influences. Whether it be a witches' spell, divine retribution, or the invisible effluvia of swamps and marshes, in all three the theme is essentially the same. Something mysterious, intangible and alien is responsible. The discovery of bacteria in the latter half of the 19th century suddenly endowed this deeply rooted cultural belief with scientific respectability. Seguin (1946) has referred to the bacterial era as 'the Golden Age for the demoniac concept of disease' and there is no doubt that the combination of these two, the half dormant cultural tradition and the power and prestige of science, served to reinforce and perpetuate the implicit assumption that all diseases were entities of some sort, attacking the individual from without and possessing some kind of shadowy independent existence of their own.

The diversity of conceptual basis to which I have referred is accompanied by a matching diversity of defining characteristics. The concept of a disease cholelithiasis arose at a time when diseases were regarded as disturbances of normal anatomy and so its defining characteristic was, and still remains, anatomical, namely the presence of stones in the gall bladder. Similarly, the concept of tuberculosis (as opposed to the much more limited concept of phthisis) developed in an era when infection by a specific micro-organism was the paradigm of disease so its defining characteristic was, and still is, the presence of the tubercle bacillus, regardless of what the patient's symptoms are or what part of his body is affected. Sometimes, however, the defining characteristic of an illness changes as knowledge of its antecedents increases, and this change may be accompanied by a subtle and sometimes unrecognized change in the population of patients to whom the diagnosis applies. Scadding (1963) quotes the example of myxoedema which was originally defined by its syndrome, but came to be defined by a disorder of function, a deficiency in the production or utilization of thyroxine. This new defining characteristic includes some patients, those with

hypopituitarism, who were not embraced by the original criterion, and excludes others, those with localized myxoedema in the absence of hypothyroidism, who were. Changes of this sort are not a problem provided they are explicit, but often they take place slowly over many years and during the transition period it is often unclear which disease is being referred to, the old or the new. A further complication resulting from diseases being defined on several different levels is that individual patients may meet the defining characteristics of two or three different diseases simultaneously, their symptomatology qualifying them for one disease, a disorder of function for a second, and an anatomical lesion for a third. A man with a bad chest, for instance, may qualify for a diagnosis of chronic bronchitis by virtue of having a chronic productive cough, for a diagnosis of asthma by being intermittently breathless as a result of variable bronchiolar narrowing, and for a diagnosis of emphysema because his pulmonary alveoli are dilated.

Such problems as these, resulting from the variable and changeable basis of the defining characteristics of different diseases, are less serious where mental illness is concerned than in most other branches of medicine because the majority of psychiatric illnesses, and all the so-called functional illnesses, are defined at the same clinical-descriptive level. The dominant conceptual model of illness in psychiatry is still the syndrome model Sydenham introduced in the 17th century, a cluster of symptoms and signs with a characteristic time course. Although in many ways this is lamentable, a cause of low reliability and witness to our ignorance, the fact that most diagnoses are defined on the same basis does at least enable psychiatry to avoid problems of this particular kind.

DISEASES AS CONCEPTS

Several times in this discussion diseases have been referred to as concepts, and reference has been made to the way in which a change in the defining characteristics of a disease may alter the population of patients embraced by the term, and even their symptoms and signs. One might add too, that it is commonplace for old concepts like Banti's syndrome and shell shock to die out, just as it is for new ones like primary aldosteronism and trisomy 21 to come into existence. In view of all this, it is well nigh impossible for us to regard illnesses as having any sort of independent material existence. To our generation it is self-evident that diseases, tuberculosis as well as schizophrenia, are nothing but man made abstractions, inventions justified only by their convenience and liable at any time to be adjusted or discarded. Our present outlook is so wholeheartedly empirical that we find it difficult to credit how an earlier generation could have talked of diseases being 'discovered' like so many golden sovereigns on a beach, or have imagined that there were a finite number of them waiting to be identified. Yet although we know these things perfectly well, we have still not rid ourselves

of all the old Platonic assumptions. Claims are still made even now that this or that syndrome is a 'disease entity', in spite of the fact that the word entity, defined in the Oxford dictionary as a 'thing that has real existence', is meaningless outside its original Platonic context. Similarly, the suggestion that patients with, for example, both schizophrenic and affective symptoms ought perhaps to be regarded as genuine interforms half way between schizophrenia and manic depressive illness still meets with expressions of shock or incredulity, as if they must really have one or other illness even if their symptoms are atypical. Part of the problem lies in the tendency of all concepts to become 'reified' and for familiarity to lead to their origins in human imagination being forgotten. The concepts of physics are as prone to this tendency as those of medicine and it may be that we are up against some of the inherent limitations of our conceptual abilities.

THE 'NON-EXISTENCE' ARGUMENT

Just at present, however, we are probably in greater danger of discarding our disease concepts completely than we are of attributing material existence to them. The pendulum has swung so far that concepts like schizophrenia and psychopathy are now in greater danger of dismissal than of reification. Szasz, Laing and others have repeatedly and gleefully asserted that 'there is no such thing as schizophrenia' or that 'schizophrenia does not exist' and been loudly applauded, at least in non-medical circles, for doing so. It is, of course, perfectly true that schizophrenia is a concept rather than an object of our senses. But the same is equally true of all other concepts, including tuberculosis and migraine, and good and evil. The fact that tuberculosis does not exist in a material sense does not prevent men dying because their lungs have been destroyed by the tubercle bacillus, nor does it relieve his headache to tell a man that there is no such thing as migraine. So it is with schizophrenia also. Even if Eugen Bleuler had never coined the term, the behavioural abnormalities, the personal experiences and the biological disadvantages which we associate with it would have remained unchanged. It is true, as Scheff (1963) and other sociologists have pointed out, that part of the disability and suffering accompanying conditions like schizophrenia is a result of the individual being labelled and treated as a schizophrenic by other people, but it is simply mischievous to suggest that schizophrenia is nothing more than the product of social or intrafamilial pressures of this kind. In fact it is equally meaningless to assert on behalf of any abstract noun or concept either that it does or that it does not exist. The only question at issue is whether it is a useful concept, and even this question has to be asked within a defined context. For the foreseeable future, the usefulness of the concept of schizophrenia is amply established by the universal occurrence of the behavioural and experiential anomalies to which the term refers, irrespective of differences in language and culture; by the biological disadvantage associated with these

anomalies, again irrespective of language or culture; by the evidence that these abnormalities are, at least in part, transmitted genetically; and by the influence on them of drugs which lack analogous effects on other people. It may well be that in time the term will lose its usefulness and pass out of use, as earlier concepts like dropsy and monomania have done, but if it does so it will be because it has been replaced by other more useful concepts, not because of any sudden realization that there is no such thing.

DIAGNOSIS

From time to time in this discussion of the concept of disease, reference has been made to the act of diagnosis and sometimes diseases have been referred to as diagnoses. The word diagnosis is derived from the Greek words *διά* (two) and *γνῶσκειν* (to know or perceive) and means literally to distinguish or to differentiate. In the medical context with which we are concerned it is used both as a verb and as a noun. The former describes the decision process by which a particular disease is attributed to a particular patient, in preference to any of the other diseases potentially applicable to him, and the latter is the decision reached, the actual illness attributed to that individual.

THE LOGIC OF CLASS MEMBERSHIP

At first sight this appears to be a straightforward exercise in determining class membership in accordance with the traditional rules of logic. In fact this is far from being the case, but it may still be helpful to describe these rules and the assumptions they involve, if only to illustrate some of the problems which diagnosis involves. The ideal abstract classes of the logician are mutually exclusive and jointly exhaustive; that is, every member of the universe possesses the defining attributes of one class of the set or array and none possesses the defining attributes of more than one. The defining attributes of the classes, and the attributes of all their members, are also 'exact', meaning that they are either present or absent, and never present in partial form. Determining the attributes of the individual and knowing the defining attributes of all the classes in the set are therefore all that is necessary to allocate the individual to the appropriate class. One might visualize, for instance, a classification of geometrical figures. Each class in the set, triangles, parallelograms, trapezoids, and so on, would be defined by the number of its sides and the angles between them. Every figure in the universe under consideration would possess the attributes of one of these classes, i.e. figures whose sides were curved or bent would automatically be excluded. Thus, counting the number of sides possessed by an individual figure and measuring the angles between these would be sufficient to determine its class membership, regardless of how many sides it possessed.

EMPIRICAL CLASSES

All real or empirical classes depart from this abstract ideal in some respects. First, they are often 'polythetic' rather than 'monothetic', meaning that they are defined not by a single necessary attribute, but by the presence of some or most of a number of attributes, none of which is mandatory for class membership. The Mammalia, for instance, are characterized by the fact that they are warm blooded, deliver their young alive, suckle them with specialized mammary glands, and possess hair and a single mandible that articulates directly with the skull. The fact that an animal lacks one or two of these characteristics does not necessarily mean that it is not a mammal, provided it possesses the others. Secondly, empirical classes are often 'inexact', meaning that their defining characteristics are capable of quantitative variation, rather than only being present or absent, so that as a result members of the domain cannot always be allocated with confidence to membership or non-membership. To pursue the same example as before, the platypus is neither cold blooded like the reptiles nor genuinely warm blooded like other mammals, but half way between the two. It maintains its body temperature above that of its environment, but is unable to keep it constant in the face of environmental fluctuations.

CLASSIFICATION OF DISEASES

Diseases share both these shortcomings of empirical classes in general. They also raise novel problems of their own. In other biological classifications it can generally be assumed quite safely that every individual belongs to one class in the array and cannot belong to two classes simultaneously, or belong to one class at one time and to another later on. The six-legged creature crawling across your desk will always belong to some insect species or other. It will not be both a beetle and a butterfly, or a beetle at one stage in its life and a butterfly at another. None of these assumptions is justified where diseases are concerned. The reason for this is that making a diagnosis is not simply a matter of allocating a human being to the appropriate class, as would be the case if he were being allocated to a species or ethnic group. Nor is it a matter of allocating an individual disease to the appropriate generic category, because a disease can only exist in the context of a diseased person. In other words, when we make a diagnosis we are not classifying people, or diseases, but sick people, and any individual person may be diseased in different ways either at different times or simultaneously. Thus, in any given instance in which the question of making a diagnosis arises, there will generally be several possibilities. It may be appropriate to make one diagnosis (e.g. chronic bronchitis), or more than one (e.g. chronic bronchitis and a duodenal ulcer) or none at all (i.e. the 'patient' may not have any disease, but simply have experienced symptoms which raised the

possibility of disease). The fact that at one point in time an individual has been confidently allocated to a particular disease category is no guarantee that on another occasion he might not have a different disease, instead or as well, or have recovered completely.

There is no easy solution to any of these problems. Certainly none of them can be eliminated or nullified by a simple change in procedure or strategy. Their most important effect is to throw the whole burden of making diagnosis a viable process onto the quality of the definitions provided for individual diseases. If these are clear cut and unambiguous, diagnosis can still be accurate and reliable, but if they are not, accuracy will deteriorate very rapidly. The reason for this is best illustrated by an analogy. Imagine a naturalist trying to identify a wild flower by comparing it with the illustrations in a flora. His task is greatly simplified by the knowledge that his flower must match one of the illustrations, and only one. If, to begin with, it seems to match two or three of them, or none at all, he is still quite likely with the help of this knowledge to end up with the right answer. But if it were possible that the flower did indeed belong to more than one of the species illustrated in his flora, or none of them, the illustrations would need to be of a very high quality for his accuracy not to be drastically reduced. The definitions we give to individual diseases correspond to the illustrations in this analogy, and need to be of high quality for the same reasons.

Mental illness is a singularly difficult territory in which to achieve this aim. Most psychiatric illnesses are still defined by their syndromes, their typical clinical features, and necessarily so because we know too little of their antecedents to define them at any other level. Many of these clinical features, like depression and anxiety, are graded traits present to varying extents in different people and at different times. Furthermore, few of them are pathognomonic of individual illnesses. In general, it is the overall pattern of symptomatology and its evolution over time that distinguishes one category of illness from another, rather than the presence of key individual symptoms. To put it in technical language, the defining characteristics of psychiatric illnesses are generally both polythetic and inexact.

OPERATIONAL DEFINITIONS

Some years ago, the philosopher Carl Hempel tactfully suggested to an audience of psychiatrists and clinical psychologists interested in problems of diagnosis and classification that they should tackle this situation by developing 'operational definitions' for all the various categories of illness in their nomenclature (Hempel, 1961). This was indeed the only advice any philosopher or scientist could have given. The term operational definition was originally coined by Bridgman (1927) who defined it as follows:

'An operational definition of a scientific term S is a stipulation to the effect

that S is to apply to all and only those cases for which performance of test operation T yields the specified outcome O'.

As Hempel himself admitted, in the context of psychiatric diagnosis the term operational has to be interpreted very liberally to include mere observation*. Really, all he is suggesting is that diagnosis S should be applied to all those, and only those, manifesting the characteristic or satisfying the criterion O, subject only to the proviso that O should be 'objective' and 'intersubjectively certifiable' and not simply something experienced intuitively or empathically by the examiner. The crux of the matter is, therefore, how to reduce a range of clinical features, many of which are quantitatively variable and none of which is usually sufficient to establish the diagnosis in question on its own, to a single objective criterion O. This is obviously a difficult and complex task. Indeed much of this book is concerned directly or indirectly with the ways in which it might be achieved. It is only appropriate at this stage to establish two general principles involved. First, individual symptoms or characteristics which are graded traits have to be converted to dichotomous variables by imposing arbitrary cut off points on them, so that the question asked is no longer 'Does the subject exhibit X?', or even 'How much X does he exhibit?', but 'Does he exhibit *this much* X?' Secondly, the traditional polythetic criterion has to be converted to a monothetic one. This can be done quite easily. Instead of saying that the typical features of disease S are A,B,C,D & E and most of these should be present before the diagnosis is made, A,B,C,D, and E must be combined algebraically, so that which combinations satisfy the criterion O and which do not is specified unambiguously. It might be stipulated for instance that any three, or any four, of the five would satisfy O, but other more complex criteria would be equally acceptable provided every possible combination was covered.

*Really it is inappropriate to talk of 'operational' definitions in a setting in which, strictly speaking, no operation is involved. 'Semantic' definition would be a better term, emphasizing that the vital element is the provision of clear cut rules of application (see p. 149). But the term 'operational' has come into widespread use and at this stage it would probably be confusing to try to change it.

3 The Issues of Reliability and Validity

RELIABILITY

Psychiatrists have always realized that their diagnoses did not invariably command the unanimous agreement of their colleagues. However it is only in the last thirty years, and at the prompting of psychologists, that the true importance of this disagreement has been appreciated, and organized attempts made to measure it. Diagnoses are of little value unless useful predictions can be made from them, and if they themselves are subject to disagreement the accuracy of these predictions will inevitably be reduced. If diagnosis A carries a 90 per cent chance of recovery and diagnosis B a 15 per cent chance, the accuracy with which the prospect of recovery can be assessed in an individual patient will largely depend on the accuracy with which A and B can be distinguished from one another. To put the matter as a general principle, the accuracy of the prognostic and therapeutic inferences derived from a diagnosis can never be higher than the accuracy with which, in any given situation, that diagnosis can itself be made; validity can never be higher than reliability and the sporadic assertions to the contrary are based on wishful thinking or sleight-of-hand.

EARLY STUDIES

The first indications that all was not well appeared in the 1930s. Masserman and Carmichael (1938) followed up a series of 100 inpatients from a university hospital in Chicago and found that a year later a 'major revision in the diagnosis' had to be made in over 40 per cent. To make matters worse, the majority of patients possessed symptoms appropriate to more than one diagnostic category both on initial examination and at follow-up. In the same year, Boisen (1938) demonstrated that the proportion of patients allocated to the different sub-categories of schizophrenia in different states, and different hospitals within a single state, varied by a factor of up to ten fold. In Massachusetts, 30 per cent of schizophrenics were catatonic, but in Illinois only 2.7 per cent. In one Illinois hospital, 76 per cent of schizophrenics were hebephrenic but in another only 11 per cent, and so on. Starting about a decade later and continuing throughout

the 1950s, numerous studies were mounted, mainly by American workers, with the specific object of testing the reliability of psychiatric diagnosis. In most of these reliability was found to be very low. In an oft-quoted study, Ash (1949) compared the diagnoses given to a series of fifty two men by three psychiatrists after a single joint interview and found that all three made the same specific diagnoses in only 20 per cent of the fifty two, and in 30 per cent all three made different diagnoses. Reviewing a series of eight studies carried out in the following decade Beck (1962) showed that, if organic cases were excluded, none of them achieved an overall agreement rate higher than 42 per cent, at least for the specific diagnosis. Nor were psychologists any more successful than psychiatrists. Goldfarb (1959) studied the diagnostic assessments of four experienced clinical psychologists in Baltimore. Their assessments were based on a Rorschach and a battery of other tests, together with a full psychiatric history, and as cases were assigned to the four quite randomly, each should have generated the same spectrum of diagnoses. In fact, the proportion of cases regarded as psychoneurotic by each varied from 4-48 per cent, and for personality disorders the range was from 22-61 per cent. Similar large differences were obtained when all four were given the same set of 100 clinical reports and test results to report on. Psychoneurotic diagnoses varied from 6-31 per cent and psychophysiological reactions from 2-16 per cent. When each of them was given the same set of clinical reports and test results again a fortnight later, 40 per cent of the subjects ended up with different diagnoses. Findings such as these produced widespread disillusionment and added further weight to the growing feeling that diagnosis was a futile exercise where mental illness was concerned. Ironically, as Zubin (1967) has pointed out, after being a major but purely academic preoccupation of psychiatrists for several decades, diagnosis fell into disrepute at the very time when the availability of a range of effective and fairly specific therapies was at last providing the exercise with some practical importance.

Unjustified pessimism

In fact, the pessimism produced by these American studies was never fully justified, for a variety of reasons. In several cases, the setting in which the study was performed weighted the scales against reliability from the beginning. In Ash's study, for instance, two thirds of the subjects were not ill at all and the diagnoses attributed to them were really only ratings of their predominant personality characteristics. Much the same is true of Goldfarb's subjects who were veterans seeking compensation or pensions for service connected disabilities. In other studies, like that of Schmidt and Fonda (1956), some or all of the diagnoses were made by residents who were in these authors' words 'young, foreign trained, inexperienced and psychiatrically naive'. In several others, like that of Hunt, Wittson and Hunt (1953), the diagnoses were made in a military

setting in which administrative considerations, like pension entitlements and criteria for discharge from the service, were likely to have been influencing admission and discharge diagnoses in different ways. And in others, like that of Stoer (1964), the study was based selectively on patients who posed diagnostic problems.

Sometimes also, the authors' conclusions were more pessimistic than their findings really justified. Pasamanick, Dinitz and Lefton (1959) compared the diagnoses given to groups of female patients randomly assigned to three different wards in the Columbus Psychiatric Institute over a two year period. Impressed by the variation in the proportion of patients allocated to different diagnostic categories on the three wards they concluded angrily: 'Clinicians, as indicated by these data, may be selectively perceiving and emphasizing only those characteristics and attributes of their patients which are relevant to their own pre-conceived system of thought... they may be so committed to a particular psychiatric school of thought, that the patient's diagnosis and treatment is largely predetermined.' However, as Kreitman (1961) has already pointed out, if one ignores the smallest group of only 48 patients the diagnostic variations between the other wards and psychiatric are all fairly modest, schizophrenia varying from 22-29 per cent, organic disorders from 7-11 per cent, psychoneuroses from 30-45 per cent, character disorders from 12-22 per cent, and so on.

However, by no means all the methodological shortcomings of these early studies were such as to lower reliability. Several must be assumed to have had quite the opposite effect. In some, of which Stoer's study is a particularly glaring example, the diagnoses being compared were not made independently of one another, in others only very broad categories of illness like psychosis and neurosis were distinguished. In Schmidt and Fonda's study, the percentage agreements obtained were inflated by statistical procedures of questionable validity. Frequently it is impossible to tell in which direction the net effect of two or more different defects or limitations in design would be likely to be. As Beck and his colleagues put it, the studies done in this era 'presented certain methodological problems which made the findings inconclusive' (Beck, Ward, Mendelson, Mock & Erbaugh, 1962).

WAYS OF MEASURING RELIABILITY

It is worth distinguishing three different ways of measuring diagnostic reliability, described by Zubin (1967) as observer agreement, frequency agreement and consistency. Measuring observer agreement is the most direct way of assessing reliability, and the most widely used. Each patient is interviewed by two or more diagnosticians either in a single joint interview or in separate interviews a few hours, or at most a few days, apart. The extent of their agreement is then usually

expressed as a percentage. Frequency agreement studies are those in which the proportion of patients allocated to individual diagnostic categories by two diagnosticians or teams of diagnosticians are compared in two series of patients whom there is good reason to regard as more or less identical. The attraction of this design is that, provided a setting is available in which this requirement of comparability is met, the study can be executed without any additional interviewing beyond that required for existing service functions. The analysis by Pasamanick, Dinitz and Lefton mentioned above is a typical example. Similar studies, with conflicting findings, have been reported by Mehlman (1952), Wilson and Meyer (1962) and Niswander, Haslerud and Weinstein (1966). Consistency studies are investigations, like that of Masserman and Carmichael, in which the diagnoses given to a group of patients at two widely separated points in time are compared. If the diagnoses are made by different people on the two occasions, as they usually are, this is really a variant of the basic observer agreement design, except that the comparatively long time interval between the two examinations increases the likelihood of disagreements being due to change in the patient, rather than to inconsistency on the part of the diagnosticians. Although the percentage agreement obtained in this setting is almost inevitably lower than in observer agreement studies, it is important for this very reason that such studies should be mounted. A diagnostic system which was unstable from week to week or month to month would be of little use however high agreement between diagnosticians might be at a single point in time.

Methodological requirements

Because of the shortcomings of the studies referred to above which he had reviewed, Beck (1962) took the trouble to stipulate what he considered to be the requirements of an adequate reliability study of observer agreement type. All the diagnosticians involved in the study should be experienced, and of comparable experience. They should agree to use a single nomenclature, preferably the most recent version of a standard or national nomenclature, and discuss any ambiguities in it before starting. The duration and setting of their interviews should be kept constant, and likewise the amount of ancillary information provided (psychological test results, information from relatives, etc.). The study should also be independent of routine assessment, to prevent diagnoses being influenced by administrative considerations. Rightly, Beck did not commit himself to a clear preference for either a single joint interview or a pair of separate interviews, though he stressed that if separate interviews were held the interval between them should be short and constant. He also pointed out that the first of these options carries with it the risk of one diagnostician guessing the trend of the other's thoughts from the questions he is asking, and the second the risk of the patient's condition changing in the interval between the two interviews.

It is also possible that the patient might react differently to the second interviewer simply because a repeat interview is a new and different situation from the first.

THE PHILADELPHIA STUDY

Beck and his colleagues in Philadelphia carried out a reliability study fulfilling these criteria on a series of 153 outpatients (Beck, Ward, Mendelson, Mock and Erbaugh, 1962). Four experienced Board-certified psychiatrists were involved and each patient was seen by two of the four in separate interviews on the same day. The allocation of psychiatrists to patients was random and each interview lasted about an hour. The nomenclature of the original (1952) edition of the *American Psychiatric Association's Diagnostic and Statistical Manual (DSM-I)* was used, augmented by some working definitions drawn up during preliminary discussions. Overall, these workers obtained 54 per cent agreement on the specific diagnosis, compared with the 15-19 per cent agreement that could have been produced by chance alone. This represents a considerable improvement on earlier studies, particularly in view of the fact that their series contained no organic patients and few psychotics. If their alternative diagnoses are also included, agreement rises to 82 per cent. But in spite of this relatively high level of agreement the authors felt bound to conclude that 'there is still a serious question as to whether the rate of agreement for the refined diagnostic categories is high enough for the purposes of research and treatment'.

A valuable and unique feature of this study was that immediately after their interviews each pair of psychiatrists compared their diagnoses and, in the forty patients over whom they disagreed, discussed why they had done so (Ward, Beck, Mendelson, Mock and Erbaugh, 1962). They considered that inconstancy on the part of the patient, either withholding information from one of the two interviewers, or giving a different account in the second interview because of insight gained in the first, was only responsible for 5 per cent of their disagreement. Inconstancy on the part of the interviewer was more important, and estimated to be the cause of 32.5 per cent of the overall disagreement. Included under this heading were differences in interviewing technique causing different items of information to be elicited by the two interviewers, and differences in the diagnostic significance attributed to symptoms which both agreed were present. They attributed most of their disagreement, however, 62.5 per cent of the total, to the inadequacies of the nomenclature they were using. They felt that it required them to make distinctions which were too fine to be practicable, did not provide adequate criteria for distinguishing related syndromes from one another, and, above all, repeatedly presented them with a forced choice between two categories (usually a psychoneurotic disorder and a personality disorder) both of whose requirements had been fulfilled.

THE CHICHESTER STUDY

At much the same time as this, Kreitman was carrying out a similar study on a series of ninety new referrals to the mental health services in Chichester (Kreitman, Sainsbury, Morrissey, Towers and Scrivener, 1961). Each patient was examined independently, first by one of three consultant psychiatrists and then, three to four days later, by one of two psychiatrists from the Medical Research Council unit. Both interviews were unstructured and were held in a variety of different settings (mainly outpatient clinics or the patient's home) and relatives were also seen wherever possible, because it was the aim of the study to stick as closely as possible to ordinary NHS working conditions. Each of the six pairs of psychiatrists involved saw a group of fifteen patients matched with the others for age, sex and social class and, as in the Philadelphia study, working definitions for a number of key terms were agreed upon beforehand. Overall diagnostic agreement, using eleven different diagnostic categories, was 63 per cent, a higher figure than that obtained by Beck and his colleagues, but probably attributable to the higher proportion of psychotic and organic patients in the series.

RELIABILITY IN RESEARCH SITUATIONS

For several years these two studies were widely regarded as an accurate indication of the best diagnostic agreement attainable by experienced psychiatrists, and so they still may be where ordinary working conditions are concerned. But for research purposes at least, considerably higher levels of agreement have now been shown to be feasible. Most of the difficulties produced by variation in interviewing technique from one psychiatrist to another can now be greatly reduced by using standardized interviewing schedules which stipulate not only which topics are to be covered, but which questions are to be asked, and in what order, and how the patients' replies are to be interpreted. Using an early version of his Present State Examination, for instance, Wing and his colleagues held joint interviews with a series of 172 patients and found that sixty nine of the seventy five patients diagnosed as schizophrenic by either rater were so diagnosed by both, a concordance rate of 92 per cent (Wing, Birley, Cooper, Graham and Isaacs, 1967).

A considerable improvement in reliability can also be obtained if all the diagnosticians involved in the study share the same training and orientation, particularly if that training has stressed the importance of observation, rather than inference, and the need for careful definition of terms. In a recent study in which audiences of psychiatrists who had all been trained at the Maudsley Hospital were asked to diagnose an unselected series of newly admitted inpatients after seeing a videotape of a brief interview with each, the average level of inter-rater agreement was 77 per cent in spite of the fact that the interviews lasted a

mere 5 minutes and no other information was provided (Kendell, 1973a). Even more impressive was the fact that the average level of agreement between the diagnoses of the videotape raters and the patients' final hospital diagnoses was still 60 per cent (compared with a chance expectation of 25 per cent), in spite of the great disparity in the amount and nature of the information on which the two sets of diagnoses were based, and the very restricted information available to the videotape audiences. As there were good grounds in this study for assuming that this figure of 60 per cent would have been significantly higher had the recorded interviews been longer, the benefits to be derived if all the raters involved in a study have been adequately trained in the same institution are clearly considerable, and well worth striving for in research settings in which diagnostic reliability is all-important.

Another simple means of improving reliability is to rank all the categories recognized, in whichever nomenclature is being used, into a hierarchy, and then stipulate that whenever the defining criteria of more than one category are satisfied the main diagnosis shall be whichever ranks higher in that hierarchy. Beck and his colleagues regarded the problem of double diagnoses, like a neurotic illness in the presence of a personality disorder, as one of the major causes of disagreement in their study, but it ought to be possible to eliminate the problem by means of a simple convention of this kind.

FACTORS INFLUENCING RELIABILITY

The whole of this discussion so far has been based on two crucial assumptions. First that it is meaningful to talk of the reliability of psychiatric diagnosis without regard to the setting, or the range of patients involved, and second, that the percentage agreement between two series of diagnoses is an adequate index of reliability. In fact, neither of these assumptions will bear close examination. The fact that they are implicit in many of the reliability studies referred to here is an important reason why this quite extensive literature provides comparatively little useful information.

Variation between diagnostic categories

There is a great deal of evidence that some individual diagnoses are considerably more reliable than others. To quote three examples: Norris (1959) studied a cohort of over 6,000 patients admitted to a number of area mental hospitals in London from a psychiatric observation ward and found that, although concordance between the mental hospital and observation ward diagnoses was 68-70 per cent for schizophrenia, manic depressive psychosis and mental deficiency, for paranoid psychoses and cerebrovascular psychosis it was only 29 per cent. Kreitman and his colleagues in their Chichester study obtained 75 per cent

agreement for organic diagnoses and 61 per cent for functional psychoses, but only 28 per cent agreement for neurotic disorders (Kreitman *et al.* 1961). Sandifer, Pettus and Quade (1964) in North Carolina, obtained 74 per cent agreement for a diagnosis of schizophrenia and 73 per cent for mental deficiency, but only 26 per cent for involutional psychotic reaction and 22 per cent for psychotic depressive reaction. [This was in a setting in which ten psychiatrists sitting as a group interviewed and diagnosed a series of ninety one patients.] In general, organic states produce higher concordance rates than functional ones, and psychoses higher concordance rates than neuroses or personality disorders. This means that global reliability is bound to vary considerably with the diagnostic composition of the population being examined, and that inpatient populations, with their relatively high proportion of psychotic patients, will generally produce higher concordance rates than outpatient series composed mainly of neurotic states.

The most obvious way of coping with this problem is to provide separate percentage agreement values for each diagnostic category or group of categories, either alongside or instead of the overall value. But even this does not really solve the problem because the reliability with which any individual category is recognized depends very much on the setting in which this recognition is taking place. For instance, a diagnosis of antisocial psychopathy might well appear to be highly reliable if three or four aggressive men with prison records were encountered in a series composed mainly of chronic schizophrenics and depressed housewives. However, if these same men were encountered in a forensic setting, the reliability of the diagnosis would probably be much lower. It is one thing to distinguish a psychopath from a schizophrenic, but quite another to distinguish him from an ordinary recidivist. Although this may be an extreme example chosen to illustrate the point, comparable situations are almost unavoidable in any reliability study. In general, the category to which any small group of patients who differ from the rest of a series are allocated will appear to be more reliable than it would in other more homogeneous settings. Conversely, the more similar all the members of a reliability study population are to one another, the harder it will be to demonstrate reliable distinctions.

The Influence of other Variables

The criteria Beck laid down for the design of an observer agreement reliability study were really those required for measuring the reliability obtainable by experienced psychiatrists under ordinary outpatient working conditions. But, as Kreitman has pointed out, we are also interested, or should be, in how reliability varies from one setting to another. Ideally, we ought to know what reliability can be achieved by experts working under optimal conditions, what by ordinary experienced psychiatrists under normal working conditions, and what by regis-

trars or residents in training. It would also be valuable to know whether diagnoses based on interviews lasting over an hour were more, or less, reliable than those lasting only twenty minutes, and how much difference it makes when information is also obtained from relatives and other ancillary sources. The fact that the overall percentage agreement achieved in any study is heavily dependent on the diagnostic composition of that particular series, and that even the values obtained for individual categories are liable to be influenced by the overall composition of the series, obviously makes it difficult to obtain clear answers to such questions. If one reliability study based on 90 minute interviews produces a higher overall percentage agreement than another based on thirty minute interviews it does not follow that diagnoses based on 90 minute interviews are more reliable than 30 minute ones, unless other factors like the competence of the raters and the diagnostic composition of the patient sample are held constant. In practice, these variables will only be held constant if both studies are part of a single research design. This implies that reliable answers to the questions we are interested in will never be provided by isolated groups of workers carrying out reliability studies of different kinds in different settings independently of one another. Doubtless, some will obtain a higher percentage agreement than others, but it will never be possible to be sure which of the many differences between them is primarily responsible for this percentage difference. Essentially this is the situation which exists at present. Since 1950 thirty or forty studies of the reliability of psychiatric diagnosis have been carried out in many different settings, but between them they convey little useful information beyond demonstrating that reliability is often very low, and generally lower for neuroses and personality disorders than for psychoses and organic states.

The limitations of percentage agreement as an index of concordance.

There are other purely statistical reasons why a bald figure for the percentage agreement between two independent series of diagnoses, or the ratio of concordant pairs of diagnoses to all pairs, conveys relatively little useful information. The first and most obvious of these is that the significance of any given percentage depends to a considerable extent on how many categories are in use. 60 per cent agreement, for instance, would mean a lot more across forty categories than it would across four. There are numerous statements in the literature to the effect that agreement was x per cent for the 'broad category' of illness and y per cent for the 'specific diagnosis'. Not only is the number of broad categories and specific diagnoses available not stated, but the meaning of these terms is also left unclear. Is schizophrenia, for example, a specific diagnosis, or does the type of schizophrenia have to be specified as well in order to qualify?

Actually, it is not sufficient just to record the number of categories available, or even the number actually used. If 70 per cent of patients are allocated to a

single category, then 49 per cent agreement could be obtained by chance alone, regardless of whether the total number of categories in use was three or thirty. Realizing this, Beck and most subsequent authors have quoted the percentage agreement that could have been produced by chance alone, given the distribution of diagnoses in their study, alongside the percentage agreement they actually obtained. What is really needed, however, is a statistic for measuring concordance which automatically allows for chance agreement and only credits agreement over and above that level. Cohen's Kappa (Cohen, 1960) fulfils this requirement and deserves to be more widely used than it is, though it is only applicable to paired comparisons and not to settings in which a single patient is diagnosed by three or more raters.

(Kappa (K) = $P_o - P_c / 1 - P_c$ where P_o is the observed agreement and P_c the chance agreement. It has a value of +1.0 if agreement is perfect, of 0 if agreement is no better than chance, and a negative value if agreement is worse than chance.)

Serious and trivial disagreements

A final problem which is ignored by most authors is that some diagnostic disagreements are more serious than others. The difference between a diagnosis of simple schizophrenia and one of a schizoid personality disorder is slight and subtle. Yet if different raters attributed these two diagnoses to the same patient, it would rank as a disagreement not only for the specific diagnosis but for the broad category of illness as well, just as if one had diagnosed simple schizophrenia and the other senile dementia. Obviously, the significance of a concordance rate of, say, 60 per cent between two raters will depend very much on whether their disagreements are mainly of the former or the latter type. Foulds (1955) tried to tackle this problem by assigning each possible combination of diagnoses a score on a 7 point agreement scale ranging from 6, perfect agreement, to 0, total disagreement. On such a scale simple schizophrenia v. schizoid personality disorder might score 3 or 4 as a partial agreement, whereas simple schizophrenia v. senile dementia would score 0. Foulds himself found, in a small scale study based on eighteen patients, that the average level of diagnostic agreement between a single pair of psychiatrists was 4. Spitzer and his colleagues in New York have discussed in some detail the inadequacies of concordance rates and contingency coefficients as indices of diagnostic agreement and pointed out that quite high percentage agreement can be obtained by chance alone, even in the presence of a statistically significant X^2 value (Spitzer, Cohen, Fleiss and Endicott, 1967). They advocate using K in preference to other more familiar statistics, and show how K can be weighted to allow for differing degrees of disagreement in the way that Foulds suggested. Weighted Kappa (K_w) is given by:

$$K_w = 1 - \frac{\sum W_{ij} P_{oij}}{\sum W_{ij} P_{cij}}$$

where W_{ij} is the weighting assigned to a given pair of diagnoses, represented by the ij cell, P_{oij} the observed proportion with that combination of diagnoses, and P_{cij} the chance proportion with that combination of diagnoses. For diagonal cells, representing identical diagnoses, $W_{ij} = 0$, so K_w , like K itself, has a value of +1.0 if agreement is perfect and of 0 if it is no better than chance (Cohen, 1968).

THE PRESENT SITUATION

The reviews of Kreitman (1961) and Beck (1962) have been referred to repeatedly in this discussion. Both gave a balanced and thoughtful account of the studies of the reliability of psychiatric diagnosis done up to that time and their conclusions were broadly in agreement with one another. Essentially, their message was that reliability was low, but not necessarily as low as some earlier studies had suggested, and that a number of ways existed in which it might be raised. Is the situation any different now, a decade later? No large scale observer agreement studies have been mounted since those Kreitman and Beck carried out themselves, perhaps because of a general realization that it was more important to improve reliability than to go on measuring unreliability. However, evidence of the depths to which diagnostic reliability is capable of sinking has continued to accumulate. Katz, Cole and Lowery (1969) presented videotapes of diagnostic interviews with two patients, admittedly chosen because they had so-called 'borderline' symptoms, to audiences of experienced psychiatrists attending meetings of the American Psychiatric Association. The forty four psychiatrists who rated the first videotape made twelve diagnoses between them, and the forty two who rated the second managed to provide fourteen different diagnoses! In each case, these diagnoses were divided more or less evenly between psychosis, neurosis and personality disorder. On the other hand, it is clear both from Katz's work and the similar videotape studies carried out by the US/UK Diagnostic Project (Kendell, Cooper, Gourlay, Copeland, Sharpe and Gurland, 1971) that other patients can command almost unanimous agreement, even between psychiatrists from very varied backgrounds. One of Katz's patients was diagnosed as a schizophrenic by all the forty psychiatrists in the audience and, of the eight videotapes shown to large groups of both American and British psychiatrists by the Diagnostic Project team, two produced almost complete agreement on a diagnosis of schizophrenia and a third unanimous agreement on a diagnosis of depression. Significantly, though, in none of these four examples was there a clear consensus for the type of schizophrenia or depression involved. This adds weight to the evidence of previous studies that it is much harder to make reliable distinctions at this level than it is to distinguish between major syndromes.

Temerlin's experiment

The most alarming study to have appeared in recent years is an experiment carried out in Oklahoma by Temerlin (1968). He played an audio tape of an assessment interview with an actor who had been carefully trained to give a convincing account of normality to audiences of psychiatrists and clinical psychologists. Before they heard the recording it was suggested to one audience that the 'patient' or client was psychotic by allowing them to overhear a high prestige figure comment that he was 'a very interesting man because he looked neurotic but actually was quite psychotic'. A second audience was similarly allowed to hear a quite different remark, 'I think this is a very rare person, a perfectly healthy man', and a third audience received no suggestion at all. After hearing the recording, the members of each audience were asked to choose a diagnosis from a list containing ten psychoses, ten neuroses and ten personality types, including 'normal or healthy personality'. All twenty members of the audience given the suggestion of health diagnosed 'normal or healthy personality'. Twelve of the twenty one members of the audience given no suggestion either way also diagnosed normality, but the other nine diagnosed various kinds of neurosis or character disorder. There were ninety five people in the audience given the suggestion that the patient was psychotic, and only eight of these decided that he was mentally healthy. Sixty of them made a diagnosis of neurosis or character disorder and twenty seven a diagnosis of psychosis, usually schizophrenia. When the different professional groups involved were considered separately, it was apparent that the psychiatrists were more suggestible than the clinical psychologists; 60 per cent of them made a diagnosis of psychosis, compared with only 28 per cent of the clinical psychologists, and none regarded the 'patient' as normal. Least suggestible of all, interestingly enough, were a group of clinical psychology students. Significantly, although everyone was asked to describe the 'behavioural basis' for his diagnosis those who diagnosed some form of illness almost invariably justified their diagnoses by inferences rather than by the descriptive statements they had been asked for. It is difficult to imagine a more vivid illustration than this of the dangers of relying on inference rather than direct observation or report, or a more telling indictment of the vague concept of psychosis employed in some circles. Even the pseudo-patients in Rosenhan's study who were diagnosed as schizophrenics by the staffs of the hospitals to which they presented themselves did at least complain of hearing voices (Rosenhan, 1973).

In spite of such findings as these, it can justifiably be claimed that the introduction of structured interviewing techniques, together with the provision of more adequate definitions for both symptoms and diagnoses, and the utilization when necessary of groups of diagnosticians who share a common training, enable psychiatric diagnoses of traditional Kraepelinian type to be made with

substantially higher reliability than that obtained by Kreitman and Beck. As the findings of Wing *et al.* (1967) and Kendell (1973a) illustrate, in research settings at least the introduction of such measures can produce acceptable levels of diagnostic concordance.

The heterogeneity of diagnostic categories

At times the diagnostic categories of psychiatry have been criticized for being heterogeneous as well as unreliable. King (1954), for example, complained that large differences both in symptomatology and in outcome were commonly found between different members of a single category and that 'the difference between two schizophrenics can be as significant as the difference between a normal and a schizophrenic'. As Zigler and Phillips (1961) have pointed out previously, such criticisms are unwarranted. It is characteristic of most biological classifications, including those of disease, to be stratified or tiered. At one level there are a small number of broad heterogeneous categories and at another a larger number of narrower and more homogeneous subcategories. The class of Mammalia is in some respects very heterogeneous but this is in no sense a shortcoming as it is subdivided into numerous genera and species which are more homogeneous. There are considerable advantages in being able to use different levels of generalization for different purposes. The same applies to categories like schizophrenia and neurosis which can also be broken up when appropriate into more homogeneous subcategories.

A more fundamental consideration, though, is that it is meaningless to state that the members of a category are either homogeneous or heterogeneous without stipulating the context. A given group of patients may be homogeneous with respect to overt behaviour, or response to phenothiazines, but highly heterogeneous with respect to mood, or previous occupation. Neither heterogeneity or homogeneity can ever be a global characteristic of any grouping; provided that group is relatively homogeneous with respect to whatever variable is under consideration at the time the grouping is justified, regardless of how diverse its members may be in other respects. It is, of course, important that a category should be relatively homogeneous in at least one chosen respect (which in the context we are concerned with may be either symptomatology, prognosis or aetiology) but this is better regarded as a question of validity than of homogeneity.

VALIDITY

It has already been stressed that reliability is a means to an end rather than an end in itself. Its importance lies in the fact that it establishes the ceiling for validity; the lower it is the lower validity necessarily becomes. The converse, of

course, is not true. Reliability can be high while validity remains trivial and in such a situation high reliability is of very limited value.

Reliability is concerned with the defining characteristics of a class, validity with the correlates of class membership. The more important correlates a given class has, over and above its defining characteristics, the greater the utility of the concept which that class represents. Psychologists have traditionally distinguished four different types of validity, or ways of establishing validity (Zubin, 1967). Briefly, these are:

1. Concurrent validity – the demonstration that independent techniques for arriving at a diagnosis both give the same diagnosis. This might be done by demonstrating that the diagnoses given to a series of patients on the basis of a diagnostic interview by a psychiatrist were the same as those assigned on the basis of responses to a self-report questionnaire like the MMPI. Alternatively, it could be done at a more fundamental level by demonstrating that groupings similar to existing diagnostic categories were produced when patients were sorted into clusters on the basis of their similarity to one another by purely statistical means.

2. Predictive validity – the demonstration that predictions derived from a diagnosis are subsequently borne out by events. These predictions may concern any aspect of prognosis; response to particular therapeutic agents or measures, the development of characteristics or behaviours which were not present at the time the diagnosis was made, mortality, the ability to leave hospital or return to work, and so on. In a sense, this is a question of homogeneity, as useful predictions can usually only be made if most members of the category behave in the same way.

3. Construct validity – the demonstration that aspects of psychopathology which can be measured objectively, like the autonomic arousal of anxiety states or the insomnia of depression, do in fact occur in the presence of diagnoses which assume their presence and not in the presence of those which assume their absence.

4. Content validity – the demonstration that the defining characteristics of a given disorder are indeed enquired into and elicited before that diagnosis is made.

THE PREDICTIVE VALIDITY OF PSYCHIATRIC DIAGNOSES

Of these four, predictive validity is by far the most important. In the last resort all diagnostic concepts stand or fall by the strength of the prognostic and therapeutic implications they embody. The ability to predict what is going to happen, and to alter this course of events if need be, have always been the main functions of medicine. Our modern commitment to elucidating aetiology is really only a strategy for furthering these practical aims rather than an end in itself.

Diagnoses like pulmonary tuberculosis and bronchial carcinoma are useful and valid not because we understand what causes them, indeed there is much about both that we do not understand, but because they enable us to predict fairly accurately what will happen to the patient, and which therapeutic measures will and will not improve that outcome.

How good then is the predictive validity of psychiatric diagnoses? Remarkably few attempts have been made to answer this question directly, though an indirect answer of a sort is implicit in every follow-up study or drug trial involving more than one diagnostic category. As a general rule, whenever two groups of patients belonging to different diagnostic categories are followed up over a period of time, whether or not they are subjected to some specified form of treatment as well, the outcome in the two differs in several statistically significant respects. On the average one group spends longer in hospital, a smaller proportion of its members recover completely, a higher proportion have to be readmitted within a year, and so on. But invariably there is a great deal of overlap between the two, perhaps 65 per cent of one group recover within 3 months but 40 per cent of the other group do likewise. The difference may be highly significant in statistical terms but only rarely is it large enough for confident predictions to be made in individual patients. To complicate the issue further, the diagnostic categories in question usually differ in a number of other respects also. They differ in age, sex and social class distribution, and perhaps in duration of illness as well. Rarely is it formally established that the modest differences in outcome and response to treatment that are observed are due to the diagnostic difference rather than to some combination of these other factors.

Perhaps the most important reason why appropriately designed studies have rarely been carried out is that much of the basic evidence for the validity of our Kraepelinian diagnostic categories accumulated long ago before people began to ask questions about reliability and validity. In the same way that no double blind controlled trial has ever been done to demonstrate the efficacy of digitalis in the treatment of congestive heart failure, because the matter was accepted as proven before controlled trials were ever thought of, so no one has ever set out to prove with matched samples of patients that manic depressive illness remits more frequently than schizophrenia, or that obsessional symptoms are more likely to persist than hysterical ones. If an entirely new classification were to be introduced its validity would undoubtedly have to be established in adequately designed and time consuming studies before it gained acceptance, just as the efficacy of new treatments has to be demonstrated by double blind trials before they supplant existing remedies. But the high suicide rate and relapse rate of manic depressive illness and the tendency to chronicity and progressive deterioration of schizophrenia are too well established, at least in the minds of psychiatrists, for it to be worthwhile at this stage designing a study to demonstrate them.

The Evidence of therapeutic trials

However, even if there is little direct evidence for, or against, the validity of time-honoured prognostic distinctions such as these, there is a great deal of information relevant to the issue of predictive validity implicit in the results of the numerous therapeutic trials that have been conducted in the last twenty years. These clearly establish the existence of several effective treatments for functional illness - ECT, the phenothiazines, the tricyclic antidepressants, lithium, and other drugs also. It is evident that each of these therapies has a limited sphere of action. ECT is most effective with manic depressive depressions, less effective with schizophrenic or manic illnesses and generally ineffective elsewhere. Phenothiazines are most effective with acute schizophrenic or manic illnesses, less effective with chronic schizophrenic or depressive illnesses and generally ineffective where neurotic syndromes are concerned. Lithium and the anti-depressant drugs show the same sort of pattern: one or two diagnostic categories where their effect is greatest, others where it is weaker but still demonstrable, and others where it is negligible. None of them has an action which is specific to one diagnostic category, and none is consistently effective in more than perhaps two thirds of the members even of that category in which its action is most powerful.

ECT can be used as a convenient illustration of this general situation. Despite the introduction of a wide range of antidepressant drugs, it is still the most effective and rapidly acting treatment for severe depressive illnesses (Medical Research Council, 1965). There is good agreement on the clinical features which correlate with response to ECT, and these correspond quite closely with the clinical features of manic depressive depression (Hobson, 1953; Roberts, 1959; Nyström, 1965). Yet in spite of this, many typical manic depressive depressions fail to respond, while other apparently unrelated syndromes, like catatonic stupor, not infrequently do so. Moreover, it is commonplace for a depressive illness to respond rapidly to ECT on one occasion but fail to do so on a subsequent one, although the patients' symptoms are just the same each time. Finally, a more accurate forecast of the likelihood of response is consistently obtained from scales based directly on symptoms than from the diagnostic category itself (Carney, Roth and Garside, 1965; Nyström, 1965). The situation is similar where each of the drugs referred to above is concerned. Even lithium, which was initially hailed as a specific treatment for the manic phase of manic depressive illness (Johnson, Gershon, Burdock, Floyd and Hekimian, 1971), now appears to be effective against both manic and excited schizoaffective illnesses, and in the latter to affect both schizophrenic and affective symptoms to much the same extent. (Prien, Caffey and Klett, 1972).

The existence of a treatment, like cobalamin for pernicious anaemia or chloroquine for malaria, which was specific for a single diagnostic category, and

almost invariably effective on members of that category, would be strong evidence for the validity of that category, but at present no such situation exists in psychiatry. The fact that each of the treatments discussed above is more effective with one category than with others tends to establish the validity of these categories, but the fact that each is effective to some extent in several categories, and inconsistent within all of them, does the reverse.

Classification by treatment response

Since the introduction of several effective psychotropic drugs in the 1950's and 60's it has often been suggested that diagnostic categories should be completely recast in terms of treatment response. Instead of schizophrenics, depressives and manics we should have phenothiazine-responders, tricyclic-responders and lithium-responders. At first glance this is an attractive idea. There is widespread dissatisfaction with our existing classification, its categories correspond imperfectly with treatment response, and one of the most important functions of classification is to determine treatment. But there are innumerable snags. Although giving a single treatment to a heterogeneous population, ECT to all depressives for example, generally provides a clear picture of the contrasting profiles of responders and non-responders, the issue becomes much more confused as soon as two or more therapeutic agents are involved.

Because the natural prognosis of most psychiatric syndromes is variable and unpredictable, conclusions about response or non-response can usually only be drawn about populations rather than individuals, and even then only in the context of a controlled trial. Drug A may be more effective than drug B in a limited group of patients, but B may have a broader field of action. C may act faster than D and so be more effective at 3 weeks, but less effective at 3 months. D may be more effective than both A and B when these are given individually, but not when they are given in combination, and so on. Above all, drug response categories are not mutually exclusive in the way that diagnostic categories are. Most tricyclic responders are also ECT responders, and some of them are phenothiazine responders as well. It is also worth remembering that drug response is rarely used as a classificatory principle in other branches of medicine, in spite of the widespread availability of effective drugs, and that because a group of patients all respond to the same drug it by no means follows that their illnesses all have the same aetiology and prognosis as well. As Hamilton once remarked, headaches, bruises and rheumatism all respond to aspirin.

Diagnosis and choice of treatment

A rather different aspect of validity to the relationship between diagnosis and the effectiveness of treatment is the relationship between diagnosis and choice

of treatment. This was studied by Bannister, Salmon and Lieberman (1964) in a population of a thousand first admissions to a London area mental hospital. After excluding all those who had been treated before admission, who refused treatment, or who had intercurrent physical illnesses, they were able to demonstrate several statistically significant relationships between the diagnosis and the treatment assigned to each. Most of these relationships, however, only held for broad diagnostic categories like schizophrenia or neurosis and nowhere was there anything approaching a one-to-one relationship. In the authors' words, 'the findings are not consistent with the notion that each particular diagnosis leads logically, or habitually, to a particular treatment. It suggests that variables other than diagnosis may be as, or more important than, diagnosis in predicated choice of treatment.' Much the same could be said of the relationship between diagnosis and the effectiveness of treatment.

EVIDENCE DERIVED BY CLUSTER ANALYSIS

An alternative, though less satisfactory, way of establishing the validity of diagnostic categories to demonstrating their predictive value is to demonstrate that they do indeed reflect independent clusters of symptoms, or circumscribed patterns of behaviour. For a remarkably long time after the framework of our present classification was laid down by Kahlbaum and Kraepelin at the beginning of the century, the existence of distinct patterns of this sort was taken for granted in most circles. It was with a sense almost of shock that Masserman and Carmichael (1938) commented on the 'mixed character' of the symptomatology of the hundred patients they studied in Chicago, and the absence of the relationship they had expected to find between current symptoms and pre-psychotic personality. Eighteen years later, Freudenberg and Robertson (1956) carried out a similar study in London and were likewise impressed and dismayed by the enormous overlap of symptoms from one diagnostic category to another, even when symptom ratings and diagnoses were both made by the same psychiatrist. However, this overlap, and the widespread occurrence of patients with symptoms appropriate to more than one diagnostic category, which are now almost universally recognized, do not necessarily imply that Kraepelin's categories are figments of the imagination. It is still possible that the symptom clusters which his disease concepts assume do exist *in re naturae*, albeit in an imperfect state. During the last twenty years elaborate statistical techniques capable, at least in principle, of demonstrating whether they do so or not have been developed and applied to psychiatric data. Using as their raw material comprehensive sets of clinical ratings derived from a large and unselected population of patients, these techniques, known generically as cluster analysis, group the patients in that population on the basis of their overall similarity into a variable number of distinct 'clusters'. Clearly, if the composition of these statistically derived clusters

corresponds closely with existing diagnostic categories this constitutes persuasive evidence for the validity of those diagnostic categories.

The principles and limitations of cluster analysis are discussed in some detail in chapter 8. All that need be said here is that many different varieties of cluster analysis have been developed, embodying different assumptions about the statistical properties of the data and utilizing different mathematical definitions of similarity. Because of these differences, different cluster analysis programs may produce different solutions to a single set of data. For example, one program may combine the patients into five clusters, while another may combine the same data into three, of which none, or only one, is recognizably the same as any of the other five. If the original material is in fact composed of fairly distinct groups of patients, this is comparatively unlikely to occur. Conversely, varying solutions of this sort are more likely to occur if there is little innate clustering tendency in the population.

This shortcoming of cluster analysis, the fact that different programs may produce widely differing solutions, was utilized by Everitt and his colleagues to demonstrate the validity of some of the traditional syndromes of psychiatry (Everitt, Gourlay and Kendell, 1971). Clinical ratings derived by structured interviewing methods from two series of 250 patients, one newly admitted to an American state hospital, the other newly admitted to an English area mental hospital, were each subjected to two quite different forms of cluster analysis. From the pool of over 700 items of information available for each patient, seventy key items were chosen and reduced by principal component analysis to ten orthogonal factors prior to the clustering procedure. In this way, four separate cluster analyses were performed on the American and British data separately by each of the two different methods. All four of these analyses produced separate clusters clearly identifiable with the manic and depressive phases of manic-depressive illness, with acute paranoid schizophrenia, and with chronic or residual schizophrenia. The manic and paranoid schizophrenic clusters were particularly clearly defined, and the chronic schizophrenic cluster the least well defined. The fact that substantially the same groups of patients were picked out by both programs from both the American and the British series is persuasive evidence that these diagnostic categories do reflect natural groupings. It is worth noting, though, that other diagnostic groups like depressive neuroses, personality disorders and alcoholism showed no tendency to form distinct clusters in any of the four analyses, in spite of being well represented in the original material.

A rather similar study has been reported by Overall (1971). A consecutive series of 350 consecutive admissions to the inpatient service of a university hospital in Texas was rated on the Factor Construct Rating Scale (FCRS), a set of seventeen 7 point scales, similar to the more well known Brief Psychiatric Rating Scale, concerned with current symptomatology and behaviour. These

ratings were subjected to a Q-type factor analysis (see chapter 8) in order to identify and delineate groups of patients with similar patterns of symptoms. In fact, the data were divided randomly into seven groups of fifty patients and the analysis carried out on each of these separately. The four clusters which were accepted from each of these seven solutions were then used as the basis for a second (higher order) analysis which produced five clusters identified by Overall as 'Depression', 'Thinking Disturbance', 'Extrapunitiveness', 'Neuroticism' and 'Agitation Excitement'. The importance of this rather complicated two-stage procedure is that only clusters which had been generated in at least four of the original seven analyses were accepted in the final solution, thereby providing some sort of guarantee that they were genuine and not mere artefacts. FCRS ratings were then obtained for a further, much larger, series of 1032 patients admitted to the same inpatient unit (Overall, Henry and Markett, 1972). Each was allocated, on the basis of a normalized vector product criterion of similarity, to one of these five clusters and the resulting allocations compared with the routine clinical diagnoses given to those same patients. There was a close relationship between the two. Most patients allocated to the depression cluster had clinical diagnoses of depression and most of those allocated to the thinking disturbance cluster had clinical diagnoses either of schizophrenia or of an organic state. Those assigned to the extrapunitiveness cluster tended to receive clinical diagnoses of psychopathy, those assigned to the neuroticism cluster to receive clinical diagnoses of psychoneurosis and those to the agitation/excitement cluster to receive clinical diagnoses of mania.

Although the authors' main aim was to validate their five empirically derived clusters, and they drew attention to several significant relationships between these and a variety of social variables and treatment assignations to that end, their results can also be regarded as a validation of the corresponding clinical categories. Although the titles they gave them were different the patient groupings Overall obtained were substantially the same as the traditional clinical groupings of depression, schizophrenia, psychopathy, psychoneurosis and mania. It could be argued, of course, that in both these cases, Overall's study and Everitt's, the close correspondence between the clusters that were obtained and traditional diagnostic groupings was produced by 'halo' effects; that in each case the raters' perception of patients' symptoms was influenced and moulded by the fact that they expected to encounter particular combinations of symptoms, and not others (Thorndike, 1920). This is an important possibility where Overall's study is concerned as his FCRS ratings were made by residents on the basis of routine unstandardized interviews conducted some time beforehand. It is a less convincing explanation of Everitt's findings as his data were derived from structured interviews carried out by a team of trained research workers.

In summary, evidence for the validity of the traditional diagnostic categories of psychiatry is of three different kinds. Evidence that different diagnostic

categories respond differently to a variety of therapeutic measures and have different short and long term prognoses, evidence that diagnostic and therapeutic assignments are related, albeit imperfectly, and statistical evidence that some major categories do reflect genuine symptom clusters. Taken as a whole, this evidence is somewhat meagre, but it is by no means non-existent at least where the functional psychoses are concerned.

VALIDITY AND USEFULNESS

The question of validity is closely bound up with that of usefulness. As we have seen, several alternative ways of establishing validity are recognized but the only one of these which also demonstrates that the categories in question have some practical utility is the demonstration that useful predictions can be derived from them. Although the validity of categories like mania, schizophrenia and depression may be demonstrated fairly convincingly by cluster analysis, their usefulness is not enhanced by such evidence. This can only be achieved by demonstrating that they carry different prognoses and respond differently to different therapeutic agents. The greater such prognostic differences are, the more useful the classification. By the same token, as long as these prognostic differences remain as weak as most are at present, the validity and usefulness of our symptom based classification will continue to be questioned.

CLASSIFICATIONS NOT BASED ON SYMPTOMS

There have, in fact, been innumerable suggestions in the last forty years that classification on the basis of symptoms should be abandoned and replaced by an entirely new classification based on data of quite a different kind. It has, for example, been suggested that psychiatric symptoms should be replaced by scores on cognitive or other psychological tests (King, 1954; Frank, 1969); by relationship patterns or styles of interpersonal behaviour (Leary and Coffey, 1955); by social role functioning and social attitudes (Clausen, 1971); by a comprehensive analysis of all behaviour in a wide range of situations (Kanfer and Saslow, 1965); and, most frequently of all, that they should be replaced by psychodynamic defence mechanisms. Almost invariably such suggestions reflect the author's own sphere of interest. With monotonous predictability, clinical psychologists suggest cognitive test results, behaviour therapists suggest total behavioural analysis and psychoanalysts suggest psychodynamic mechanisms. But almost never has any serious attempt been made to develop, test and use an alternative classification on any of these bases. The suggestion is made and the advantages of the innovation emphasized, but it is left to others to implement it. The would-be innovator either lacks the courage of his convictions, or is daunted by the enormity of the task which he has set himself. One may suspect

that a classification based on psychodynamic defence mechanisms would be hamstrung by the low reliability common to all inferential judgements, that one based on cognitive test results would yield even fewer useful prognostic distinctions than we have now, and that one based on an analysis of the patient's total behavioural repertoire would simply prove impracticable, but one can do no more than suspect these things because classifications on these bases have never progressed beyond the stage of advocacy. The plain reason why psychiatrists continue to use their traditional symptom-based classification in spite of its low reliability and limited validity is that, as things stand, there is no alternative.

The question of whether the relationship between one patient and another is better expressed by a set of categories or by a set of dimensions has not been raised here, partly because it is discussed in detail in chapter 9 and partly because it is an independent issue. Whether symptoms, or cognitive test results, or psychodynamic mechanisms, or some combination of all three, is chosen as the most appropriate basis for classification both options, a typology or a dimensional system, are still available.

4 Diagnosis as a Practical Decision-Making Process

Although a great deal has been said and written about the logical status of diagnoses, and their reliability has been measured and questioned many times, surprisingly little interest has been taken in the practical aspects of diagnosis as a decision-making process. In the 1950's when diagnosis and assessment procedures in general were the main professional concern of clinical psychologists, a considerable literature was generated by them on such matters as the contribution of individual cognitive or projective tests to various tests batteries, the difference between the diagnostic assessments of experienced and inexperienced psychologists, and the difference between assessments made by those who administered the tests and those made by others who were simply given the results to interpret. But as the diagnoses with which this work was concerned were almost exclusively based on test results – cognitive, projective, preceptual or motor – often without the patient even being seen by the diagnostician, the relevance of their findings to the traditional medical method of making a diagnosis by taking a history and examining the mental state is questionable. Even in well designed studies yielding unambiguous results, like that of Kostlan (1954), it is difficult for psychiatrists to draw valid inferences about their own diagnostic behaviour, beyond learning that psychometric test results in general, and projective tests in particular, are of little real value in most situations.

In recent years a few studies have been published in which psychiatrists and other physicians have, for the first time, examined their own diagnostic activities in specially designed experimental situations. The results of these experiments are interesting and instructive, and worth describing in some detail in spite of the doubts that must exist about the validity of drawing conclusions about diagnostic behaviour in real life from behaviour in essentially artificial situations.

The situation in which psychiatric diagnoses are most commonly made is one in which a clinician, armed with a variable amount of background information, like the patient's age and occupation and source of referral, holds a free-ranging

interview with the patient, lasting anything from twenty minutes to an hour or more. In this interview he seeks to establish a diagnosis by asking the patient first about his current symptoms and difficulties and then about an ever widening circle of other experiences and events, past and present. If one considers this paradigm situation for a few minutes a number of questions immediately spring to mind:

QUESTIONS TO BE ANSWERED

1. What kinds of information does the clinician use to arrive at a diagnosis, and are some types of information consistently more valuable than others?
2. Can one distinguish distinct styles of gathering information and of reasoning from this information, and are some of these superior to others?
3. What is the role of experience and other personal characteristics in determining the variation in accuracy and economy with which clinicians arrive at diagnoses?
4. How long does it usually take an interviewer to reach a confident diagnosis, and at what stage in a diagnostic interview are the crucial decisions usually made?
5. Are some diagnoses consistently easier to make than others, and if so, why?

A convenient way of describing the little we know about the way in which psychiatric diagnoses are made, and the results of the experimental studies referred to above, is to consider each of these five questions in turn.

The type of information used

It has been demonstrated in several different settings that there is no simple relationship between a patient's symptoms and the diagnosis assigned to him even when the same clinicians are responsible for both. (Wittenborn, Holzberg and Simon, 1953; Freudenberg and Robertson, 1956; Nathan, Gould, Zare and Roth, 1969). Although in theory diagnosis is based on symptomatology, and consistent, albeit modest, symptomatic differences can be demonstrated between populations drawn from different diagnostic categories, attempts to define the exact relationship between the two in ordinary clinical practice have usually ended in exasperation or bafflement. As both a cause and an effect of this imperfect relationship between symptoms and diagnosis there is considerable variation in the type and quantity of the information different psychiatrists try to obtain before making a diagnosis. Some concentrate mainly on the patient's current mental state or some particular aspect of this, his overt behaviour in the interview for instance, or the degree of insight he shows into the origins of his symptoms. Others are much more concerned with his personality as a whole,

his constitutional endowment, his success in coping with life's demands and his habitual ways of reacting to stress. Others focus on the situation in which his symptoms developed and the reasons why this was so stressful to him. Many of these differences can be related to different conceptions of the nature of mental illness, or to the demands of the setting in which the psychiatrist is operating, so it is not always meaningful to ask which approach is best, or to attempt to choose experimentally between one approach and another. Even allowing for such problems, the fact remains that most psychiatrists attempt to collect quite a lot of information of varying kinds before making a diagnosis, and do so without any real knowledge of which types of information are most valuable, and why. The content of the diagnostic interview is determined by a blend of practical experience and inherited tradition and depends on these, *faute de mieux*, for its justification.

The results of an ingenious study by Gauron and Dickinson (1966a) in Iowa suggest that, as one might expect, there is considerable variation in the diagnostic usefulness of the various items of information psychiatrists seek to acquire in diagnostic interviews. The results of this study also suggest that psychiatrists have little idea which are the most useful items and which the least. What Gauron and Dickinson did was this. They analysed the case histories of three patients and artificially divided each of them into thirty six different information categories - childhood history, projective test results, reason for referral, previous personality, and so on. A group of twelve psychiatrists were then allowed to ask for these units of information, one at a time, in whatever order they liked. After receiving each new piece of information they were required to suggest one or more diagnoses, together with an estimate of the probability of its being correct. This process continued until they either had all thirty six units of information, or were confident in their final diagnosis. At the end the twelve participants were asked to list in order, for each of the three patients, the five pieces of information they felt had been most important in determining their final diagnosis. By recording which items of information had first suggested his three final diagnoses to each participant, and which had increased the probability ratings of these at each stage thereafter, the authors were able to calculate an actual order of importance for the thirty six items, to compare with the perceived order. Not unexpectedly, 'reason for referral' came first by a wide margin in both lists, but thereafter there was little or no relationship between the two and the rank order correlation between them was too small to be statistically significant.

Traditionally, the mental state and the history have always been regarded as the two basic components of the diagnostic interview, the latter sometimes being obtained wholly or in part from a relative. Which of these is, or should be, more important and how often relatives or other informants provide crucial information not obtainable from the patient are both questions which have rarely been posed. Simon, Gurland, Fleiss and Sharpe (1971) analysed how frequently,